

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 April 2002 (25.04.2002)

PCT

(10) International Publication Number
WO 02/032925 A2

- (51) International Patent Classification⁷: **C07K**
- (21) International Application Number: PCT/US01/32233
- (22) International Filing Date: 16 October 2001 (16.10.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/688,566 16 October 2000 (16.10.2000) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 09/688,566 (CIP)
Filed on 16 October 2000 (16.10.2000)
- (71) Applicant (for all designated States except US): **PHYLOS, INC.** [US/US]; 128 Spring Street, Lexington, MA 02421 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **LIPOVSEK, Dasa** [SI/US]; 45 Sunset Road, Cambridge, MA 02138 (US). **WAGNER, Richard, W.** [US/US]; 1007 Lowell Road, Concord, MA 01742 (US). **KUIMELIS, Robert, G.** [US/US]; 21 Malbert Road, Brighton, MA 02135 (US).
- (74) Agent: **ELBING, Karen, L.**; Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110-2214 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:
6 February 2003

(15) Information about Correction:
see PCT Gazette No. 06/2003 of 6 February 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/032925 A2

(54) Title: PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS AND OTHER BINDING PROTEINS

(57) Abstract: Disclosed herein are proteins that include an immunoglobulin fold and that can be used as scaffolds. Also disclosed herein are nucleic acids encoding such proteins and the use of such proteins in diagnostic methods and in methods for evolving novel compound-binding species and their ligands.

PROTEIN SCAFFOLDS FOR ANTIBODY MIMICS
AND OTHER BINDING PROTEINS

5

Background of the Invention

This invention relates to protein scaffolds useful, for example, for the generation of products having novel binding characteristics.

Proteins having relatively defined three-dimensional structures, commonly referred to as protein scaffolds; may be used as reagents for the design of engineered products. These scaffolds typically contain one or more regions which are amenable to specific or random sequence variation, and such sequence randomization is often carried out to produce libraries of proteins from which desired products may be selected. One particular area in which such scaffolds are useful is the field of antibody design.

A number of previous approaches to the manipulation of the mammalian immune system to obtain reagents or drugs have been attempted. These have included injecting animals with antigens of interest to obtain mixtures of polyclonal antibodies reactive against specific antigens, production of monoclonal antibodies in hybridoma cell culture (Koehler and Milstein, Nature 256:495, 1975), modification of existing monoclonal antibodies to obtain new or optimized recognition properties, creation of novel antibody fragments with desirable binding characteristics, and randomization of single chain antibodies (created by connecting the variable regions of the heavy and light chains of antibody molecules with a flexible peptide linker) followed by selection for antigen binding by phage display (Clackson et al., Nature 352:624, 1991).

In addition, several non-immunoglobulin protein scaffolds have been proposed for obtaining proteins with novel binding properties. For example, a "minibody" scaffold, which is related to the immunoglobulin fold, has been designed by deleting three beta strands from a heavy chain variable domain of a monoclonal antibody (Tramontano et al., J. Mol. Recognit, 7:9, 1994). This

protein includes 61 residues and can be used to present two hypervariable loops. These two loops have been randomized and products selected for antigen binding, but thus far the framework appears to have somewhat limited utility due to solubility problems. Another framework used to display loops has been
5 tendamistat, a 74 residue, six-strand beta sheet sandwich held together by two disulfide bonds (McConnell and Hoess, J. Mol. Biol. 250:460, 1995). This scaffold includes three loops, but, to date, only two of these loops have been examined for randomization potential.

Other proteins have been tested as frameworks and have been used to
10 display randomized residues on alpha helical surfaces (Nord et al., Nat. Biotechnol. 15:772, 1997; Nord et al., Protein Eng. 8:601, 1995), loops between alpha helices in alpha helix bundles (Ku and Schultz, Proc. Natl. Acad. Sci. USA 92:6552, 1995), and loops constrained by disulfide bridges, such as those of the small protease inhibitors (Markland et al., Biochemistry 35:8045, 1996;
15 Markland et al., Biochemistry 35:8058, 1996; Rottgen and Collins, Gene 164:243, 1995; Wang et al., J. Biol. Chem. 270:12250, 1995).

Summary of the Invention

The present invention provides a new family of proteins capable of evolving to bind any compound of interest. These proteins, which generally
20 make use of a scaffold derived from a fibronectin type III (Fn3) or Fn3-like domain, function in a manner characteristic of natural or engineered antibodies (that is, polyclonal, monoclonal, or single-chain antibodies) and, in addition, possess structural advantages. Specifically, the structure of these antibody mimics has been designed for optimal folding, stability, and solubility, even
25 under conditions that normally lead to the loss of structure and function in antibodies.

These antibody mimics may be utilized for the purpose of designing proteins which are capable of binding to virtually any compound (for example, any protein) of interest. For example, the ¹⁰Fn3-based molecules described

herein may be used as scaffolds which are subjected to directed evolution to form a population with one or more randomized Fn3 loops that are analogous by position and structure to the complementarity-determining regions (CDRs) of an antibody variable region, and/or to randomize Fn3's other three solvent exposed loops. Such a directed evolution approach results in the production of antibody-like molecules with high affinities for antigens of interest. In addition, the scaffolds described herein may be used to display defined exposed loops (for example, loops previously randomized and selected on the basis of antigen binding) in order to direct the evolution of molecules that bind to such introduced loops. A selection of this type may be carried out to identify recognition molecules for any individual CDR-like loop or, alternatively, for the recognition of two or all three CDR-like loops combined into a non-linear epitope.

Accordingly, in a first aspect, the present invention features randomized or mutated scaffold proteins. In particular, the invention features a non-antibody protein including a domain having an immunoglobulin-like fold, the non-antibody protein deriving from a reference protein by having a mutated amino acid sequence, wherein the non-antibody protein binds with a K_d at least as tight as $1\ \mu\text{M}$ to a compound that is not bound as tightly by the reference protein.

In addition, the invention features a non-antibody protein deriving from a scaffold protein including a domain having an immunoglobulin-like fold, wherein the amino acid sequence of the domain in the derived protein is more than 50% identical to the amino acid sequence of the domain in the scaffold protein.

In yet another embodiment, the invention features a protein that includes a fibronectin type III domain having at least one randomized loop, the protein being characterized by the ability of the Fn3 domain to bind to a compound that is not bound by the corresponding naturally-occurring Fn3 domain.

In various preferred embodiments, any of these proteins of the invention bind to their target compounds with a K_d at least as tight as 500 nM, preferably,

with a K_d at least as tight as 100nM or 10nM, and, more preferably, with a K_d at least as tight as 1 nM, 500 pM, 100pM, or even 20 pM. The protein preferably contains one, two, or three mutated loops and at least one of the loops, and preferably two or all three of the loops, contributes to the binding of
5 the protein to the compound. Additionally, the reference protein preferably lacks disulfide bonds, and the derivative protein may have at least one disulfide bond.

With respect to certain embodiments, the domain having an immunoglobulin-like fold preferably has a molecular mass less than 10kD or
10 greater than 7.5 kD, and, more preferably, has a molecular mass between 7.5-10 kD. The proteins of the invention may be monomers under physiological conditions or may be multimers, for example, dimers. In other preferred embodiments, the reference protein used to derive a mutated protein of the invention is a naturally-occurring mammalian protein (for example, a human
15 protein); and the domain having an immunoglobulin-like fold is mutated and includes up to 50%, and preferably up to 34%, mutated amino acids as compared to the immunoglobulin-like fold of the reference protein. In addition, the domain having the immunoglobulin-like fold preferably consists of approximately 50-150 amino acids, and more preferably consists of
20 approximately 50 amino acids.

Derivative proteins of the invention may be derived from any appropriate reference protein including, but not limited to, the preferred proteins, fibronectin or a fibronectin dimer, tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-receptor, cytokine receptor, glycosidase inhibitor, antibiotic
25 chromoprotein, myelin membrane adhesion molecule PO, CD8, CD4, CD2, class I MHC, T-cell antigen receptor, CD1, C2 and I-set domains of VCAM-1, I-set immunoglobulin domain of myosin-binding protein C, I-set immunoglobulin domain of myosin-binding protein H, I-set immunoglobulin domain of telokin, NCAM, twitchin, neuroglian, growth hormone receptor, erythropoietin receptor,
30 prolactin receptor, interferon-gamma receptor, β -galactosidase/glucuronidase, β -

glucuronidase, transglutaminase, T-cell antigen receptor, superoxide dismutase, tissue factor domain, cytochrome F, green fluorescent protein, GroEL, and thaumatin.

In further preferred embodiments of Fn3 domain-containing proteins, the
5 fibronectin type III domain is a mammalian (for example, a human) fibronectin type III domain; and the protein includes the tenth module of the fibronectin type III (¹⁰Fn3) domain. In such proteins, compound binding is preferably mediated by either one, two, or three ¹⁰Fn3 loops. In other preferred
10 embodiments, the second (DE) loop of ¹⁰Fn3 may be extended in length relative to the naturally-occurring module, or the ¹⁰Fn3 may lack an integrin-binding motif. In these molecules, the integrin-binding motif may be replaced by an amino acid sequence in which a polar amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the
15 integrin-binding motif; alternatively, one preferred sequence is serine-glycine-glutamate. In another preferred embodiment, the fibronectin type III domain-containing proteins of the invention lack disulfide bonds.

Any of the proteins of the invention (for example, the fibronectin type III domain-containing proteins) may be formulated as part of a fusion protein. If
20 the fusion protein is to be used for compound binding or compound binding selections, the fusion protein includes a heterologous protein that does not itself bind to the compound of interest. The heterologous protein may, for example, be an antibody or antibody domain (such as an immunoglobulin F_c domain), a complement protein, a toxin protein, or an albumin protein. In addition, any of
25 the proteins of the invention (for example, the fibronectin type III domain proteins) may be covalently bound to a nucleic acid (for example, an RNA), and the nucleic acid may encode the protein. Moreover, the protein may be a multimer, or, particularly if it lacks an integrin-binding motif, it may be formulated in a physiologically-acceptable carrier.

The present invention also features proteins that include a fibronectin type
30 III domain having at least one mutation in a β -sheet sequence. Again, these

proteins are characterized by their ability to bind to compounds that are not bound or are not bound as tightly by the corresponding naturally-occurring fibronectin domain.

Any of the proteins of the invention may be immobilized on a solid support (for example, a bead or chip), and these proteins may be arranged in any configuration on the solid support, including an array.

In a related aspect, the invention further features nucleic acids encoding any of the proteins of the invention. In preferred embodiments, the nucleic acid is DNA or RNA.

10 In another related aspect, the invention also features a method for generating a protein which includes a fibronectin type III domain and which is pharmaceutically acceptable to a mammal, involving removing the integrin-binding domain of said fibronectin type III domain. This method may be applied to any of the fibronectin type III domain-containing proteins described
15 above and is particularly useful for generating proteins for human therapeutic applications. The invention also features such fibronectin type III domain-containing proteins which lack integrin-binding domains.

In yet another related aspect, the invention features methods of obtaining derivative non-antibody proteins which bind to compounds of interest. One such
20 method involves: (a) providing a non-antibody scaffold protein including an immunoglobulin-like fold, wherein the scaffold protein does not bind to the compound with a K_d as tight as $1 \mu M$; (b) generating mutated derivatives of the non-antibody scaffold protein, thereby producing a library of mutated proteins; (c) contacting the library with the compound; (d) selecting from the library at
25 least one derivative protein which binds to the compound with a K_d at least as tight as $1 \mu M$; and (e) optionally repeating steps (b) - (d) substituting for the non-antibody scaffold protein in repeated step (b) the product from the previous step (d). This technique may also be carried out with any of the proteins of the invention (for example, any of the fibronectin type III domain-containing
30 proteins).

In yet other related aspects, the invention features screening methods which may be used to obtain or evolve randomized or mutated proteins of the invention capable of binding to compounds of interest, or to obtain or evolve compounds (for example, proteins) capable of binding to a particular protein containing a randomized or mutated motif. In addition, the invention features screening procedures which combine these two methods, in any order, to obtain either compounds or proteins of interest.

In particular, a first screening method, useful for the isolation or identification of randomized or mutated proteins of interest, involves: (a) contacting a compound of interest with a candidate protein, the candidate protein being a derivative non-antibody protein including a domain having an immunoglobulin-like fold, the non-antibody protein deriving from a reference protein by having a mutated amino acid sequence wherein the non-antibody protein binds with a K_d at least as tight as $1 \mu M$ to a compound that is not bound as tightly by the reference protein, wherein the contacting is carried out under conditions that allow compound-protein complex formation; and (b) obtaining, from the complex, the derivative protein that binds to the compound. This general technique may also be carried out with a fibronectin type III domain protein having at least one randomized or mutated loop.

The second screening method is for isolating or identifying a compound which binds to a protein of the invention. This method begins with a non-antibody protein including a domain having an immunoglobulin-like fold and deriving from a reference protein by having a mutated amino acid sequence, wherein the non-antibody protein binds with a K_d at least as tight as $1 \mu M$ to a compound that is not bound as tightly by the reference protein. This derivative protein is then contacted with a candidate compound, wherein the contacting is carried out under conditions that allow compound-protein complex formation, and the compound which binds to the derivative protein is obtained from the complex. Again, this general technique may be carried out with any protein of the invention, for example, a protein with a mutated fibronectin type III domain.

In addition, the invention features diagnostic methods which employ the proteins of the invention (for example, fibronectin type III scaffold proteins and their derivatives). Such diagnostic methods may be carried out on a sample (for example, a biological sample) to detect one analyte or to simultaneously detect
5 many different analytes in the sample. The method may employ any of the scaffold molecules described herein. Preferably, the method involves (a) contacting the sample with a protein of the invention that binds to the compound analyte, the contacting being carried out under conditions that allow compound-protein complex formation; and (b) detecting the complex, and therefore the
10 compound in the sample. In addition, this method may be used to quantitate, as well as detect, compound levels in a sample.

In preferred embodiments of any of the selection or diagnostic methods described herein, the protein of the invention binds to its target compound with a K_d at least as tight as 1 μ M or 500 nM, preferably, with a K_d at least as tight as
15 100nM or 10nM, and, more preferably, with a K_d at least as tight as 1 nM, 500 pM, 100pM, or even 20 pM. The protein preferably contains one, two, or three mutated loops and at least one of the loops, and preferably two or all three of the loops contributes to the binding of the protein to the compound. Additionally, the reference protein preferably lacks disulfide bonds, and the derivative protein
20 may have at least one disulfide bond.

With respect to certain embodiments of the methods, the domain having an immunoglobulin-like fold preferably has a molecular mass less than 10kD or greater than 7.5 kD, and, more preferably, has a molecular mass between 7.5-10 kD. The proteins of the invention may be monomers under physiological
25 conditions or may be multimers, for example, dimers. In other preferred embodiments, the reference protein used to derive a mutated protein of the invention is a naturally-occurring mammalian protein (for example, a human protein); and the domain having an immunoglobulin-like fold is mutated and includes up to 50%, and preferably **up** to 34%, mutated amino acids as compared
30 to the immunoglobulin-like fold of the reference protein. In addition, the

domain having an immunoglobulin-like fold preferably consists of approximately 50-150 amino acids, and more preferably consists of approximately 50 amino acids.

- Derivative proteins used in the methods of the invention may be derived
- 5 from any appropriate reference protein including, but not limited to, the preferred proteins, fibronectin or a fibronectin dimer, tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-receptor, cytokine receptor, glycosidase inhibitor, antibiotic chromoprotein, myelin membrane adhesion molecule PO, CD8, CD4, CD2, class I MHC, T-cell antigen receptor, CD1, C2 and I-set
 - 10 domains of VCAM-1, I-set immunoglobulin domain of myosin-binding protein C, I-set immunoglobulin domain of myosin-binding protein H, I-set immunoglobulin domain of telokin, NCAM, twitchin, neuroglian, growth hormone receptor, erythropoietin receptor, prolactin receptor, interferon-gamma receptor, β -galactosidase/glucuronidase, β -glucuronidase, transglutaminase, T-
 - 15 cell antigen receptor, superoxide dismutase, tissue factor domain, cytochrome F, green fluorescent protein, GroEL, and thaumatin.

- In addition, the steps of the selection methods described herein may be repeated with further mutation or randomization being carried out between cycles. For example, for the methods involving a protein having a mutated or
- 20 randomized fibronectin type III domain, at least one loop of the fibronectin type III domain of the protein obtained in step (b) may be mutated and steps (a) and (b) repeated using the further randomized protein, or the compound obtained in step (b) may be modified and steps (a) and (b) repeated using the further modified compound. In these methods, the compound is preferably a protein,
 - 25 and the fibronectin type III domain is preferably a mammalian (for example, a human) fibronectin type III domain. In other preferred embodiments, the protein includes the tenth module of the fibronectin type III domain ($^{10}\text{Fn3}$), and binding is mediated by one, two, or three $^{10}\text{Fn3}$ loops. In addition, the second (DE) loop of $^{10}\text{Fn3}$ may be extended in length relative to the naturally-occurring module, or
 - 30 $^{10}\text{Fn3}$ may lack an integrin-binding motif. Again, as described above, the

integrin-binding motif may be replaced by an amino acid sequence in which a basic amino acid-neutral amino acid-acidic amino acid sequence (in the N-terminal to C-terminal direction) replaces the integrin-binding motif; alternatively, one preferred replacement sequence is serine-glycine-glutamate.

- 5 The selection and diagnostic methods described herein may be carried out using any of the proteins of the invention (for example, a fibronectin type III domain-containing protein). In addition, any of these proteins may be formulated as part of a fusion protein with a heterologous protein (for example, an antibody or antibody domain (including an immunoglobulin F_c domain) that
- 10 does not itself bind the compound of interest, or a complement protein, toxin protein, or albumin protein). In addition, selections and diagnostic methods may be carried out using the proteins of the invention (for example, the fibronectin type III domain proteins) covalently bound to nucleic acids (for example, RNAs or any nucleic acid which encodes the protein). Moreover, the selections and
- 15 diagnostic methods may be carried out using these proteins (for example, the fibronectin domain-containing proteins) as monomers or as multimers, such as dimers.

Preferably, the selections and diagnostic methods involve the immobilization of the binding target on a solid support. Preferred solid supports

20 include columns (for example, affinity columns, such as agarose-based affinity columns), microchips, or beads. Alternatively, the proteins (for example, the Fn3 domain-containing proteins) may be immobilized and contacted with one or more potential binding targets.

For the diagnostic methods, the compound is often a protein, but may also

25 be any other analyte in a sample. Detection may be accomplished by any standard technique including, without limitation, radiography, fluorescence detection, mass spectroscopy, or surface plasmon resonance.

In a final aspect, the invention features a non-antibody protein that binds tumor necrosis factor- α (TNF- α) with a K_d at least as tight as 1 μ M, the protein

having a sequence that is less than 20% identical to TNF-a receptor (for example, a naturally-occurring TNF-a receptor, such as a mammalian or human TNF-a receptor).

In preferred embodiments, this protein includes a mutated fibronectin type
5 III domain and the protein is mutated in the DE, BC, and FG loops. Preferably, the mutated FG loop is the same length as the wild-type FG loop. In other preferred embodiments, the protein includes an immunoglobulin-like fold (preferably, having a molecular mass less than 10 kD, greater than 7.5 kD, or between 7.5-10 kD) that consists of approximately 50-150 amino acids, and
10 preferably, approximately 50 amino acids.

The TNF-a binders according to the invention bind TNF-a with a Kd at least as tight as 1 μ M, preferably, at least as tight as 500 nM, 100 nM, or 10 nM, more preferably, at least as tight as 1 nM or 500 pM, and, most preferably, at least as tight as 100 pM or even 20 pM. Preferably, these proteins contain one,
15 two, or three mutated loops, and at least one, and preferably two or all three of the loops, contribute to the binding of the non-antibody protein to TNF-a. In other preferred embodiments, the non-antibody protein has at least one disulfide bond, and the non-antibody protein is a monomer or dimer under physiological conditions.

20 The TNF-a binders may be immobilized on a solid support (for example, a chip or bead), and may be part of an array. In addition, any of the TNF-a binders may be joined to a heterologous protein (for example, a heterologous protein that is an antibody or an antibody domain that does not bind TNF-a, an immunoglobulin F_c domain, a complement protein, or an albumin protein).

25 If desired, the protein may include a mutated fibronectin type III domain (for example, one derived from a human fibronectin type III domain, such as a mutated tenth module of the fibronectin type III domain (¹⁰F_n3)). In addition, the protein may lack an ¹⁰F_n3 integrin-binding motif. TNF-a binders preferably include a non-naturally occurring sequence in a loop of ¹⁰F_n3 (for example, the

loop sequence PW(A/G), and may include a non-naturally occurring sequence in a β -sheet of ¹⁰F_n3. Particularly preferred TNF- α binders of the invention are shown in Figure 25 (SEQ ID NOS: 34-140).

In addition, in related aspects, the invention features nucleic acids
5 encoding any of the TNF- α binding proteins of the invention, as well as a loop structure on any protein that includes any one of the amino acid sequences of Figure 25 (SEQ ID NOS: 34-140).

As used herein, by "non-antibody protein" is meant a protein that is not produced by the B cells of a mammal either naturally or following immunization
10 of a mammal. This term also excludes antibody fragments of more than 100 amino acids, preferably, more than 80 amino acids, and, most preferably, more than 50 amino acids in length.

By "immunoglobulin-like fold" is meant a protein domain of between about 80-150 amino acid residues that includes two layers of antiparallel beta-
15 sheets, and in which the flat, hydrophobic faces of the two beta-sheets are packed against each other. Proteins according to the invention may include several immunoglobulin-like folds covalently bound or associated non-covalently into larger structures.

By "scaffold" is meant a protein used to select or design a protein
20 framework with specific and favorable properties, such as binding. When designing proteins from the scaffold, amino acid residues that are important for the framework's favorable properties are retained, while others residues may be varied. Such a scaffold has less than 50% of the amino acid residues that vary between protein derivatives having different properties and greater than or equal
25 to 50% of the residues that are constant between such derivatives. Most commonly, these constant residues confer the same overall three-dimensional fold to all the variant domains, regardless of their properties.

By "fibronectin type III domain" is meant a domain having 7 or 8 beta strands which are distributed between two beta sheets, which themselves pack
30 against each other to form the core of the protein, and further containing loops

which connect the beta strands to each other and are solvent exposed. There are at least three such loops at each edge of the beta sheet sandwich, where the edge is the boundary of the protein perpendicular to the direction of the beta strands. Preferably, a fibronectin type III domain includes a sequence which exhibits at least 30% amino acid identity, and preferably at least 50% amino acid identity, to the sequence encoding the structure of the ¹⁰F_n3 domain referred to as "1ttg" (ID = "1ttg" (one ttg)) available from the RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Base. Sequence identity referred to in this definition is determined by the Homology program, available from Molecular Simulation (San Diego, CA). The invention further includes polymers of ¹⁰F_n3-related molecules, which are an extension of the use of the monomer structure, whether or not the subunits of the polyprotein are identical.

By "naturally occurring" is meant any protein that is encoded by a living organism.

By "randomized" or "mutated" is meant including one or more amino acid alterations relative to a template sequence. By "randomizing" or "mutating" is meant the process of introducing, into a sequence, such an amino acid alteration. Randomization or mutation may be accomplished through intentional, blind, or spontaneous sequence variation, generally of a nucleic acid coding sequence, and may occur by any technique, for example, PCR, error-prone PCR, or chemical DNA synthesis. By a "corresponding, non-mutated protein" is meant a protein that is identical in sequence, except for the introduced amino acid mutations.

By a "protein" is meant any sequence of two or more amino acids, regardless of length, post-translation modification, or function. "Protein" and "peptide" are used interchangeably herein.

By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term
5 includes, without limitation, DNA, RNA, and PNA.

By "pharmaceutically acceptable" is meant a compound or protein that may be administered to an animal (for example, a mammal) without significant adverse medical consequences.

By "physiologically acceptable carrier" is meant a carrier which does not
10 have a significant detrimental impact on the treated host and which retains the therapeutic properties of the compound with which it is administered. One exemplary physiologically acceptable carrier is physiological saline. Other physiologically acceptable carriers and their formulations are known to one skilled in the art and are described, for example, in Remington's Pharmaceutical
15 Sciences, (18" edition), ed. A. Gennaro, 1990, Mack Publishing Company, Easton, PA, incorporated herein by reference.

By a "fusion protein" is meant a protein that includes (i) a scaffold protein of the invention joined to (ii) a second, different (i.e., "heterologous") protein. "Fusion proteins" are distinguished from "nucleic acid-protein fusions" and
20 "RNA-protein fusions" in that a "fusion protein" is composed entirely of amino acids, while both a "nucleic acid-protein fusion" and an "RNA-protein fusion" include a stretch of nucleic acids (the nucleic acid or **RNA** component) joined to a stretch of amino acids (the protein component).

By "selecting" is meant substantially partitioning a molecule from other
25 molecules in a population. As used herein, a "selecting" step provides at least a 2-fold, preferably, at least a 30-fold, more preferably, at least a 100-fold, and, most preferably, at least a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps
30 may be combined in a given approach.

By "binding partner," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired compound (for example, protein) of interest. Examples of binding partners include, without limitation, members of antigen/antibody pairs, protein/inhibitor pairs, 5 receptor/ligand pairs (for example cell surface receptor/ligand pairs, such as hormone receptor/peptide hormone pairs), enzyme/substrate pairs (for example, kinase/substrate pairs), lectin/carbohydrate pairs, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pairs, RNA/protein pairs, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as 10 any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of another molecule (for example, a compound or protein).

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, agarose or 15 sepharose), microchip (for example, silicon, silicon-glass, or gold chip), or membrane (for example, an inorganic membrane, nitrocellulose, or the membrane of a liposome or vesicle) to which an antibody mimic or an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which 20 an antibody mimic or an affinity complex may be embedded (for example, through a receptor or channel).

The present invention provides a number of advantages. For example, as described in more detail below, the present antibody mimics exhibit improved biophysical properties, such as stability under reducing conditions and solubility 25 at high concentrations. In addition, these molecules may be readily expressed and folded in prokaryotic systems, such as E. coli, in eukaryotic systems, such as yeast, and in in vitro translation systems, such as the rabbit reticulocyte lysate system. Moreover, these molecules are extremely amenable to affinity maturation techniques involving multiple cycles of selection, including in vitro 30 selection using RNA-protein fusion technology (Roberts and Szostak, Proc.

Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al. WO98/31700), phage display (see, for example, Smith and Petrenko, Chem. Rev. 97:3 17, 1997), and yeast display systems (see, for example, Boder and Wittrup, Nature Biotech. 15:553, 1997).

- 5 Other features and advantages of the present invention will be apparent from the following detailed description thereof, and from the claims.

Brief Description of the Drawings

FIGURE 1 is a photograph showing a comparison between the structures of antibody heavy chain variable regions from camel (dark blue) and llama (light
10 blue), in each of two orientations.

FIGURE 2 is a photograph showing a comparison between the structures of the camel antibody heavy chain variable region (dark blue), the llama antibody heavy chain variable region (light blue), and a fibronectin type III module number 10 (¹⁰F_n3) (yellow).

- 15 FIGURE 3 is a photograph showing a fibronectin type III module number 10 (¹⁰F_n3), with the loops corresponding to the antigen-binding loops in IgG heavy chains highlighted in red.

FIGURE 4 is a graph illustrating a sequence alignment between a fibronectin type III protein domain and related protein domains.

- 20 FIGURE 5 is a photograph showing the structural similarities between a ¹⁰F_n3 domain and 15 related proteins, including fibronectins, tenascins, collagens, and undulin. In this photograph, the regions are labeled as follows: constant, dark blue; conserved, light blue; neutral, white; variable, red; and RGD integrin-binding motif (variable), yellow.

- 25 FIGURE 6 is a photograph showing space filling models of fibronectin III modules 9 and 10, in each of two different orientations. The two modules and the integrin binding loop (RGD) are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

FIGURE 7 is a photograph showing space filling models of fibronectin III modules 7-10, in each of three different orientations. The four modules are labeled. In this figure, blue indicates positively charged residues, red indicates negatively charged residues, and white indicates uncharged residues.

5 FIGURE 8 is a photograph illustrating the formation, under different salt conditions, of RNA-protein fusions which include fibronectin type III domains.

FIGURE 9 is a series of photographs illustrating the selection of fibronectin type III domain-containing RNA-protein fusions, as measured by PCR signal analysis.

10 FIGURE 10 is a graph illustrating an increase in the percent TNF- α binding during the selections described herein, as well as a comparison between RNA-protein fusion and free protein selections.

FIGURE 11 is a series of schematic representations showing IgG, $^{10}\text{Fn3}$, $\text{Fn-CH}_1\text{-CH}_2\text{-CH}_3$, and $\text{Fn-CH}_2\text{-CH}_3$ (clockwise from top left).

15 FIGURE 12 is a photograph showing a molecular model of $\text{Fn-CH}_1\text{-CH}_2\text{-CH}$, based on known three-dimensional structures of IgG (X-ray crystallography) and $^{10}\text{Fn3}$ (NMR and X-ray crystallography).

FIGURE 13 is a graph showing the time course of an exemplary $^{10}\text{Fn3}$ -based nucleic acid-protein fusion selection of TNF- α binders. The
20 proportion of nucleic acid-protein fusion pool (open diamonds) and free protein pool (open circles) that bound to TNF- α -Sephadex, and the proportion of free protein pool (full circles) that bound to underivatized Sephadex, are shown.

FIGURES 14 and 15 are graphs illustrating TNF- α binding by TNF- α Fn-
25 binders. In particular, these figures show mass spectra data obtained from a $^{10}\text{Fn3}$ fusion chip and non-fusion chip, respectively.

FIGURES 16 and 17 are the phosphorimage and fluorescence scan, respectively, of an $^{10}\text{Fn3}$ array, illustrating TNF- α binding.

FIGURE 18 is a graph showing an alignment of the primary sequences of the llama V_H domain and the wild-type human $^{10}\text{Fn3}$ domain. Homologous
30 residues between the two sequences are indicated. The $^{10}\text{Fn3}$ residues outside

the randomized loops that were found to have mutated in approximately 45% of the selected clones are marked with arrows under the wild-type ¹⁰F_n3 sequence and with the letter that identifies the selected residue.

FIGURE 19 shows schematic representations of the llama V_H domain and the wild-type human ¹⁰F_n3 domain. The locations of the mutated framework residues are indicated.

FIGURE 20 is a graph illustrating the efficiency and specificity of binding of a free-protein pool translated from the original library (RO) and after ten rounds of selection with TNF-α (R10). Protein pool binding to underivatized
10 Sepharose, to TNF-α-Sepharose, to IL-1α-Sepharose, and to IL-13-Sepharose is compared.

FIGURE 21 is a series of IgG-like scaffolds for the display of up to three loops.

FIGURE 22 is a series of IgG-like scaffolds for the display of up to four,
15 or even six, loops.

FIGURE 23 is a series of scaffolds, unrelated to IgG, for the display of loop structures.

FIGURES 24A-24D are photographic and graphic illustrations demonstrating the specific capture of a target (TNF-α) by a mimic immobilized
20 on a solid surface.

FIGURE 25 is a graph listing exemplary TNF-α binders (SEQ ID NOS: 33-140) according to the invention.

Detailed Description

The novel antibody mimics described herein have been designed to be
25 superior both to antibody-derived fragments and to non-antibody frameworks, for example, those frameworks cited above.

The major advantage of these antibody mimics over antibody fragments is structural. These antibody mimics are derived from whole, stable, and soluble structural scaffolds. For example, the F_n3 scaffold is found in the human body.

Consequently, they exhibit better folding and thermostability properties than antibody fragments, whose creation involves the removal of parts of the antibody native fold, often exposing amino acid residues that, in an intact antibody, would be buried in a hydrophobic environment, such as an interface
5 between variable and constant domains. Exposure of such hydrophobic residues to solvent increases the likelihood of aggregation of the antibody fragments.

In addition, the scaffolds described herein have no disulfide bonds, which have been reported to retard or prevent proper folding of antibody fragments under certain conditions. Since the present scaffolds do not rely on disulfides
10 for native fold stability, they are stable under reducing conditions, unlike antibodies and their fragments which unravel upon disulfide bond reduction.

Moreover, these scaffolds provide the functional advantages of antibody molecules. In particular, despite the fact that the ¹⁰F_n3 module is not an immunoglobulin, its overall fold is close to that of the variable region of the IgG
15 heavy chain (Figure 2), making it possible to display the three fibronectin loops analogous to CDRs in relative orientations similar to those of native antibodies. Because of this structure, the present antibody mimics possess antigen binding properties that are similar in nature and affinity to those of antibodies, and a loop randomization and shuffling strategy may be employed in vitro that is similar to
20 the process of affinity maturation of antibodies in vivo.

There are now described below exemplary scaffolds, for example, fibronectin-based scaffolds, and their use for identifying, selecting, and evolving novel binding proteins as well as their target ligands. These examples are provided for the purpose of illustrating, and not limiting, the invention.

25 ¹⁰F_n3 Structural Motif

Preferred antibody mimics of the present invention are based on the structure of a fibronectin module of type III (F_n3), a common domain found in mammalian blood and structural proteins. This domain occurs more than 400 times in the protein sequence database and has been estimated to occur in 2% of

the proteins sequenced to date, including fibronectins, tenascin, intracellular cytoskeletal proteins, and prokaryotic enzymes (Bork and Doolittle, Proc. Natl. Acad. Sci. USA 89:8990, 1992; Bork et al., Nature Biotech. 15:553, 1997; Meinke et al., J. Bacteriol. 175:1910, 1993; Watanabe et al., J. Biol. Chem. 5 265:15659, 1990). A particular scaffold is the tenth module of human Fn3 (¹⁰Fn3), which comprises 94 amino acid residues. The overall fold of this domain is closely related to that of the smallest functional antibody fragment, the variable region of the heavy chain, which comprises the entire antigen recognition unit in camel and llama IgG (Figure 1, 2). The major differences 10 between camel and llama domains and the ¹⁰Fn3 domain are that (i) ¹⁰Fn3 has fewer beta strands (seven vs. nine) and (ii) the two beta sheets packed against each other are connected by a disulfide bridge in the camel and llama domains, but not in ¹⁰Fn3.

The three loops of ¹⁰Fn3 corresponding to the antigen-binding loops of the 15 IgG heavy chain run between amino acid residues 21-31 (BC), 51-56 (DE), and 76-88 (FG) (Figure 3). The length of the BC and DE loop, 10 and 6 residues, respectively, fall within the narrow range of the corresponding antigen-recognition loops found in antibody heavy chains, that is, 7-10 and 4-8 residues, respectively. Accordingly, once randomized and selected for high 20 antigen affinity, these two loops may make contacts with antigens equivalent to the contacts of the corresponding loops in antibodies.

In contrast, the FG loop of ¹⁰Fn3 is 12 residues long, whereas the corresponding loop in antibody heavy chains ranges from 4-28 residues. To optimize antigen binding, therefore, the length of the FG loop of ¹⁰Fn3 is 25 preferably randomized in length as well as in sequence to cover the CDR3 range of 4-28 residues to obtain the greatest possible flexibility and affinity in antigen binding. Indeed, in general, the lengths as well as the sequences of the CDR-like loops of the antibody mimics may be randomized during in vitro or in vivo affinity maturation (as described in more detail below).

The tenth human fibronectin type III domain, ¹⁰F_n3, refolds rapidly even at low temperature; its backbone conformation has been recovered within 1 second at 5°C. Thermodynamic stability of ¹⁰F_n3 is high ($\Delta G_U = 24 \text{ kJ/mol} = 5.7 \text{ kcal/mol}$), correlating with its high melting temperature of 110°C.

- 5 One of the physiological roles of ¹⁰F_n3 is as a subunit of fibronectin, a glycoprotein that exists in a soluble form in body fluids and in an insoluble form in the extracellular matrix (Dickinson et al., J. Mol. Biol. 236:1079, 1994). A fibronectin monomer of 220-250 kD contains 12 type I modules, two type II modules, and 17 fibronectin type III modules (Potts and Campbell, Curr. Opin. Cell Biol. 6:648, 1994). Different type III modules are involved in the binding of fibronectin to integrins, heparin, and chondroitin sulfate. ¹⁰F_n3 was found to mediate cell adhesion through an integrin-binding Arg-Gly-Asp (RGD) motif on one of its exposed loops. Similar RGD motifs have been shown to be involved in integrin binding by other proteins, such as fibrinogen, von
- 10 Wellebrand factor, and vitronectin (Hynes et al., Cell 69: 11, 1992). No other matrix- or cell-binding roles have been described for ¹⁰F_n3.

- The observation that ¹⁰F_n3 has only slightly more adhesive activity than a short peptide containing RGD is consistent with the conclusion that the cell-binding activity of ¹⁰F_n3 is localized in the RGD peptide rather than
- 20 distributed throughout the ¹⁰F_n3 structure (Baron et al., Biochemistry 31:2068, 1992). The fact that ¹⁰F_n3 without the RGD motif is unlikely to bind to other plasma proteins or extracellular matrix makes ¹⁰F_n3 a useful scaffold to replace antibodies. In addition, the presence of ¹⁰F_n3 in natural fibrinogen in the bloodstream suggests that ¹⁰F_n3 itself is unlikely to be immunogenic in the
- 25 organism of origin.

- In addition, we have determined that the ¹⁰F_n3 framework possesses exposed loop sequences tolerant of randomization, facilitating the generation of diverse pools of antibody mimics. This determination was made by examining the flexibility of the ¹⁰F_n3 sequence. In particular, the human ¹⁰F_n3 sequence
- 30 was aligned with the sequences of fibronectins from other sources as well as

sequences of related proteins (Figure 4), and the results of this alignment were mapped onto the three-dimensional structure of the human ¹⁰Fn3 domain (Figure 5). This alignment revealed that the majority of conserved residues are found in the core of the beta sheet sandwich, whereas the highly variable residues are

5 located along the edges of the beta sheets, including the N- and C-termini, on the solvent-accessible faces of both beta sheets, and on three solvent-accessible loops that serve as the hypervariable loops for affinity maturation of the antibody mimics. In view of these results, the randomization of these three loops are unlikely to have an adverse effect on the overall fold or stability of the

10 ¹⁰Fn3 framework itself.

For the human ¹⁰Fn3 sequence, this analysis indicates that, at a minimum, amino acids 1-9, 44-50, 61-54, 82-94 (edges of beta sheets); 19, 21, 30-46 (even), 79-65 (odd) (solvent-accessible faces of both beta sheets); 21-31, 51-56, 76-88 (CDR-like solvent-accessible loops); and 14-16 and 36-45 (other

15 solvent-accessible loops and beta turns) may be randomized to evolve new or improved compound-binding proteins. In addition, as discussed above, alterations in the lengths of one or more solvent exposed loops may also be included in such directed evolution methods.

Alternatively, changes in the β -sheet sequences may also be used to evolve

20 new proteins. These mutations change the scaffold and thereby indirectly alter loop structure(s). If this approach is taken, mutations should not saturate the sequence, but rather few mutations should be introduced. Preferably, no more than between 3-20 changes should be introduced to the β -sheet sequences by this approach.

25 Sequence variation may be introduced by any technique including, for example, mutagenesis by Taq polymerase (Tindall and Kunkel, Biochemistry 27:6008 (1988)), fragment recombination, or a combination thereof. Similarly, an increase of the structural diversity of libraries, for example, by varying the length as well as the sequence of the CDR-like loops, or by structural redesign

based on the advantageous framework mutations found in selected pools, may be used to introduce further improvements in antibody mimic scaffolds.

Antibody Mimic Fusions

- The antibody mimics described herein may be fused to other protein domains. For example, these mimics may be integrated with the human immune response by fusing the constant region of an IgG (F_c) with an antibody mimic, such as an ¹⁰F_n3 module, preferably through the C-terminus of ¹⁰F_n3. The F_c in such a ¹⁰F_n3-F_c fusion molecule activates the complement component of the immune response and increases the therapeutic value of the antibody mimic.
- 10 Similarly, a fusion between an antibody mimic, such as ¹⁰F_n3, and a complement protein, such as C1q, may be used to target cells, and a fusion between an antibody mimic, such as ¹⁰F_n3, and a toxin may be used to specifically destroy cells that carry a particular antigen. In addition, an antibody scaffold, such as ¹⁰F_n3, in any form may be fused with albumin to increase its half-life in the
- 15 bloodstream and its tissue penetration. Any of these fusions may be generated by standard techniques, for example, by expression of the fusion protein from a recombinant fusion gene constructed using publically available gene sequences.

Scaffold Multimers

- In addition to monomers, any of the scaffold constructs described herein
- 20 may be generated as dimers or multimers of antibody mimics as a means to increase the valency and thus the avidity of antigen binding. Such multimers may be generated through covalent binding. For example, individual ¹⁰F_n3 modules may be bound by imitating the natural ⁸F_n3-⁹F_n3-¹⁰F_n3 C-to-N-terminus binding or by imitating antibody dimers that are held together
- 25 through their constant regions. A ¹⁰F_n3-Fc construct may be exploited to design dimers of the general scheme of ¹⁰F_n3-Fc::Fc-¹⁰F_n3. The bonds engineered into

the Fc::Fc interface may be covalent or non-covalent. In addition, dimerizing or multimerizing partners other than Fc can be used in hybrids, such as ¹⁰F_n3 hybrids, to create such higher order structures.

- In particular examples, covalently bonded multimers may be generated by
- 5 constructing fusion genes that encode the multimer or, alternatively, by engineering codons for cysteine residues into monomer sequences and allowing disulfide bond formation to occur between the expression products. Non-covalently bonded multimers may also be generated by a variety of techniques. These include the introduction, into monomer sequences, of codons
- 10 corresponding to positively and/or negatively charged residues and allowing interactions between these residues in the expression products (and therefore between the monomers) to occur. This approach may be simplified by taking advantage of charged residues naturally present in a monomer subunit, for example, the negatively charged residues of fibronectin. Another means for
- 15 generating non-covalently bonded antibody mimics is to introduce, into the monomer gene (for example, at the amino- or carboxy-termini), the coding sequences for proteins or protein domains known to interact. Such proteins or protein domains include coil-coil motifs, leucine zipper motifs, and any of the numerous protein subunits (or fragments thereof) known to direct formation of
- 20 dimers or higher order multimers.

Fibronectin-Like Molecules

- Although ¹⁰F_n3 represents a preferred scaffold for the generation of antibody mimics, other molecules may be substituted for ¹⁰F_n3 in the molecules described herein. These include, without limitation, human fibronectin modules
- 25 ¹F_n3-⁹F_n3 and ¹¹F_n3-¹⁷F_n3 as well as related F_n3 modules from non-human animals and prokaryotes. In addition, F_n3 modules from other proteins with sequence homology to ¹⁰F_n3, such as tenascins and undulins, may also be used. Other exemplary scaffolds having immunoglobulin-like folds (but with sequences that are unrelated to the V_H domain) are shown in Figure 21 and

include N-cadherin, ICAM-2, titin, GCSF receptor, cytokine receptor, glycosidase inhibitor, E-cadherin, and antibiotic chromoprotein. Yet further domains with related structures may be derived from myelin membrane adhesion molecule PO, CD8, CD4, CD2, class I MHC, T-cell antigen receptor, CD1, C2
5 and I-set domains of VCAM-1, I-set immunoglobulin domain of myosin-binding protein C, I-set immunoglobulin domain of myosin-binding protein H, I-set immunoglobulin domain of telokin, telikin, NCAM, twitchin, neuroglian, growth hormone receptor, erythropoietin receptor, prolactin receptor, GC-SF receptor, interferon-gamma receptor, β -galactosidase/glucuronidase, β -
10 glucuronidase, and transglutaminase. Alternatively, any other protein that includes one or more immunoglobulin-like folds may be utilized. Such proteins may be identified, for example, using the program SCOP (Murzin et al., J. Mol. Biol. 247:536 (1995); Lo Conte et al., Nucleic Acids Res. 25:257 (2000).

Generally, any molecule that exhibits a structural relatedness to the V_H
15 domain (as identified, for example, using the computer program above) may be utilized as an antibody mimic. Such molecules may, like fibronectin, include three loops at the N-terminal pole of the molecule and three loops at the C-terminal pole, each of which may be randomized to create diverse libraries; alternatively, larger domains may be utilized, having larger numbers of loops, as
20 long as a number of such surface randomizable loops are positioned closely enough in space that they can participate in antigen binding. Figure 22 shows examples of useful domains having more than three loops positioned close to each other. These examples include T-cell antigen receptor and superoxide dismutase, which each have four loops that can be randomized; and an Fn3
25 dimer, tissue factor domains, and cytokine receptor domains, each of which have three sets of two similar domains where three randomizable loops are part of the two domains (bringing the total number of loops to six).

In yet another alternative, any protein having variable loops positioned close enough in space may be utilized for candidate binding protein production.
30 For example, large proteins having spatially related, solvent accessible loops

may be used, even if unrelated structurally to an immunoglobulin-like fold. Exemplary proteins include, without limitation, cytochrome F, green fluorescent protein, GroEL, and thaumatin (Figure 23). The loops displayed by these proteins may be randomized and superior binders selected from a randomized library as described herein. Because of their size, molecules may be obtained that exhibit an antigen binding surface considerably larger than that found in an antibody-antigen interaction. Other useful scaffolds of this type may also be identified using the program SCOP (Murzin et al., J. Mol. Biol. 247:536 (1995)) to browse among candidate proteins having numerous loops, particularly loops positioned among parallel beta sheets or a number of alpha-helices.

Modules from different organisms and parent proteins may be most appropriate for different applications. For example, in designing an antibody mimic, it may be most desirable to generate that protein from a fibronectin or fibronectin-like molecule native to the organism for which a therapeutic is intended. In contrast, the organism of origin is less important or even irrelevant for antibody mimics that are to be used for in vitro applications, such as diagnostics, or as research reagents.

For any of these molecules, libraries may be generated and used to select binding proteins by any of the methods described herein.

20 Directed Evolution of Scaffold-Based Binding Proteins

The antibody mimics described herein may be used in any technique for evolving new or improved binding proteins. In one particular example, the target of binding is immobilized on a solid support, such as a column resin or microtiter plate well, and the target contacted with a library of candidate scaffold-based binding proteins. Such a library may consist of antibody mimic clones, such as ¹⁰Fn3 clones constructed from the wild type ¹⁰Fn3 scaffold through randomization of the sequence and/or the length of the ¹⁰Fn3 CDR-like loops. If desired, this library may be an RNA-protein fusion library generated, for example, by the techniques described in Szostak et al., U.S.S.N. 09/007,005

and 09/247,190; Szostak et al., WO98/3 1700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Alternatively, it may be a DNA-protein library (for example, as described in Lohse, DNA-Protein Fusions and Uses Thereof, U.S.S.N. 60/1 10,549, U.S.S.N. 09/459,190, and WO 5 00/32823). The fusion library is incubated with the immobilized target, the support is washed to remove non-specific binders, and the tightest binders are eluted under very stringent conditions and subjected to PCR to recover the sequence information or to create a new library of binders which may be used to repeat the selection process, with or without further mutagenesis of the 10 sequence. A number of rounds of selection may be performed until binders of sufficient affinity for the antigen are obtained.

In one particular example, the ¹⁰F_n3 scaffold may be used as the selection target. For example, if a protein is required that binds a specific peptide sequence presented in a ten residue loop, a single ¹⁰F_n3 clone is constructed in 15 which one of its loops has been set to the length of ten and to the desired sequence. The new clone is expressed *in vivo* and purified, and then immobilized on a solid support. An RNA-protein fusion library based on an appropriate scaffold is then allowed to interact with the support, which is then washed, and desired molecules eluted and re-selected as described above.

20 Similarly, the scaffolds described herein, for example, the ¹⁰F_n3 scaffold, may be used to find natural proteins that interact with the peptide sequence displayed by the scaffold, for example, in an ¹⁰F_n3 loop. The scaffold protein, such as the ¹⁰F_n3 protein, is immobilized as described above, and an RNA-protein fusion library is screened for binders to the displayed loop. The 25 binders are enriched through multiple rounds of selection and identified by DNA sequencing.

In addition, in the above approaches, although RNA-protein libraries represent exemplary libraries for directed evolution, any type of scaffold-based library may be used in the selection methods of the invention.

Use

- The antibody mimics described herein may be evolved to bind any antigen of interest. These proteins have thermodynamic properties superior to those of natural antibodies and can be evolved rapidly in vitro. Accordingly, these
- 5 antibody mimics may be employed in place of antibodies in all areas in which antibodies are used, including in the research, therapeutic, and diagnostic fields. In addition, because these scaffolds possess solubility and stability properties superior to antibodies, the antibody mimics described herein may also be used under conditions which would destroy or inactivate antibody molecules.
- 10 Finally, because the scaffolds of the present invention may be evolved to bind virtually any compound, these molecules provide completely novel binding proteins which also find use in the research, diagnostic, and therapeutic areas.

Experimental Results

- Exemplary scaffold molecules described above were generated and tested,
- 15 for example, in selection protocols, as follows.

Library construction

- A complex library was constructed from three fragments, each of which contained one randomized area corresponding to a CDR-like loop. The randomized residues are indicated in Figure 18 as underlined sequences,
- 20 specifically, residues 23-29 of the ¹⁰F_n3 BC loop (corresponding to CDR-H1 of the llama V_H); residues 52-55 of the ¹⁰F_n3 DE loop (corresponding to CDR-H2 of the llama V_H); and residues 78-87 of the ¹⁰F_n3 FG loop (corresponding to CDR-H3 of the llama V). The fragments were named BC, DE, and FG based on the names of the CDR-H-like loops contained within them; in addition to
- 25 ¹⁰F_n3 and a randomized sequence, each of the fragments contained stretches encoding an N-terminal His₆ domain or a C-terminal FLAG peptide tag. At each junction between two fragments (i.e., between the BC and DE fragments or between the DE and FG fragments), each DNA fragment contained recognition

sequences for the *Eco*RI Type IIS restriction endonuclease. This restriction enzyme allowed the splicing together of adjacent fragments while removing all foreign, non-¹⁰F_n3, sequences. It also allowed for a recombination-like mixing of the three ¹⁰F_n3 fragments between cycles of mutagenesis and selection.

- 5 The wild-type, human ¹⁰F_n3 gene was cloned from a human liver library (Maxim Biotech, South San Francisco, CA) using the primers Hu5PCR-NdeI 5'CATATGGTTTCTGATGTTCCGAGG3'; SEQ ID NO: 28) and Hu3PCR-EcoRI (5'GAATTCCTATGTTCCGTAATTAATGGAAATTG3'; SEQ ID NO: 29). Three different libraries were constructed from the wild-type
- 10 segments obtained by the PCR of the ¹⁰F_n3 clone and from randomized segments obtained by oligonucleotide synthesis. The BC_r-DE_r-FG_r library was obtained by randomizing the selected residues in BC, DE, and FG loops; the BC_r-DE_{wt}-FG_r library was obtained by randomizing the selected residues in BC and FG loops, leaving the DE loop sequence wild-type; and the BC_{wt}-DE_{wt}-FG_r
- 15 library was obtained by randomizing the selected residues in the FG loop only.

The BC_r, DE_r, and FG_r fragments were made synthetically. Each fragment was assembled from two overlapping oligonucleotides, which were first annealed, then extended to form the double-stranded DNA form of the fragment. The oligonucleotides that were used to construct and process the three fragments

20 are listed below; the "Top" and "Bottom" species for each fragment are the oligonucleotides that contained the entire ¹⁰F_n3 encoding sequence. In these oligonucleotides designations, "N" indicates A, T, C, or G; and "S" indicates C or G.

HfnLbcTop (His):

- 25 5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT
ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT GTT
CCG AGG GAC CTG GAA GTT GTT GCT GCG ACC CCC ACC AGC-3'
(SEQ ID NO: 1)

HfnLbcTop (an alternative N-terminus):

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT
ACA ATT ACA ATG GTT TCT GAT GTT CCG AGG GAC CTG GAA GTT
GTT GCT GCG ACC CCC ACC AGC-3' (SEQ ID NO: 2)

5 HFnLBCBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC
TGT TTC TCC GTA AGT GAT CCT GTA ATA TCT (SNN), CCA GCT GAT
CAG TAG GCT GGT GGG GGT CGC AGC -3' (SEQ ID NO: 3)

HFnBC3'-flag8:

10 5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CCC
TGT TTC TCC GTA AGT GAT CC-3' (SEQ ID NO: 4)

HFnLDETop:

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT
ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC ACA GGA
15 GGA AAT AGC CCT GTC C-3' (SEQ ID NO: 5)

HFnLDEBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT
ATA ATC AAC TCC AGG TTT AAG GCC GCT GAT GGT AGC TGT
(SNN), AGG CAC AGT GAA CTC CTG GAC AGG GCT ATT TCC TCC
20 TGT -3' (SEQ ID NO: 6)

HFnDE3'-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC GCT CTT CGT
ATA ATC AAC TCC AGG TTT AAG G-3' (SEQ ID NO: 7)

HF_nLFGTop:

5'- GG AAT TCC TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT
ACA ATT ACA ATG CAT CAC CAT CAC CAT CAC CTC TTC TAT ACC
ATC ACT GTG TAT GCT GTC-3' (SEQ ID NO: 8)

5 HF_nLFGBot-flag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA
ATT AAT GGA AAT TGG (SNN),, AGT GAC AGC ATA CAC AGT GAT
GGT ATA -3' (SEQ ID NO: 9)

HF_nFG3'-flag8:

10 5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC TGT TCG GTA
ATT AAT GGA AAT TGG -3' (SEQ ID NO: 10)

T7T_{mv} (introduces T7 promoter and TMV untranslated region needed for in
vitro translation):

5'- GCG TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA
15 ATT ACA-3' (SEQ ID NO: 11)

ASAflag8:

5'-AGC GGA TGC CTT GTC GTC GTC GTC CTT GTA GTC-3' (SEQ ID NO:
12)

Unispl-s (spint oligonucleotide used to ligate mRNA to the
20 puromycin-containing linker, described by Roberts et al, 1997, supra):

5'-TTTTTTTTTTNAGCGGATGC-3' (SEQ ID NO: 13)

A18---2PEG (DNA-puromycin linker):

5'-(A)₁₈(PEG)₂CCP_{ur} (SEQ ID NO: 14)

The oligonucleotide pair BC_{Top} and BC_{Bot-flag8} was used to construct the fragment which contains the randomized BC loop; the pair DE_{Top} and DE_{Bot-flag8} was used to construct the fragment which contains the randomized DE loop; the pair BC_{Top} and DE_{3-Flag8} was used to PCR-amplify the BC_n - DE_{wt} fragments; and
5 the pair FG_{Top} and FG_{Bot-Flag8} was used to construct the fragment which contains the randomized FG loop. The pairs of oligonucleotides (500 pmol of each) were annealed in 100 μ L of 10 mM Tris 7.5, 50 mM NaCl for 10 minutes at 85°C, followed by a slow (0.5-1 hour) cooling to room temperature. The annealed fragments with single-stranded overhangs were then extended using 100 U
10 Klenow (New England Biolabs, Beverly, MA) for each 100 μ L aliquot of annealed oligos, and the buffer made of 838.5 μ l H₂O, 9 μ l 1 M Tris 7.5, 5 μ l 1M MgCl₂, 20 μ l 10mM dNTPs, and 7.5 μ l 1M DTT. The extension reactions proceeded for 1 hour at 25°C.

In order to reduce the frequency of stop codons introduced by the random
15 sequences, the randomized residues were encoded by (NNS)_n, where N stands for any nucleotide and S for an equimolar mixture of C and G; only one of the three stop codons (TAG) conforms to the NNS restriction. In addition to the sequence encoding ¹⁰F_n3, the gene fragments contained the 5' Tobacco Mosaic Virus (TMV) untranslated region and the T7 promoter, as well as the sequences
20 encoding a 5' hexahistidine protein purification tag and a 3' FLAG epitope purification tag. In addition, as noted above, Ear I restriction endonuclease recognition sites were engineered into the overlaps between adjacent fragments in order to facilitate the assembly of the three fragments.

Next, each of the double-stranded fragments was transformed into an
25 RNA-protein fusion (PROfusion™) using the technique developed by Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/3 1700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302. Briefly, the fragments were transcribed using an Ambion in vitro transcription kit, T7-MEGAshortscript™ (Ambion, Austin, TX), and the
30 resulting mRNA was gel-purified and ligated to a 5'-phosphorylated

DNA-puromycin linker, preferably, 5' dA₁₈PEG₂dCdCPur) using DNA ligase (Promega, Madison, WI); the mRNA was aligned with the DNA linker using a DNA splint oligonucleotide (5' TTTTTTTTTNAGCGGATGC 3'; SEQ ID NO: 30) as described in Szostak (*supra*). The mRNA-DNA-puromycin molecule was
5 then translated using the Ambion rabbit reticulocyte lysate-based translation kit in the presence of ³⁵S-methionine. The resulting mRNA-DNA-puromycin-protein fusion was purified using Oligo(dT) cellulose, (Type 7, Amersham Pharmacia, Piscataway, NJ) and a complementary DNA strand was synthesized using reverse transcriptase (Superscript™II, Gibco, Life Technologies,
10 Rockville, MD) and the RT primers described above (Unisplint-S or flagASA), following the manufacturer's instructions (preferably, a two-minute annealing at 70°C and a 40 minute reaction at 42°C).

The RNA-protein fusion with annealing cDNA obtained for each fragment was next purified on the resin appropriate to its peptide purification tag, i.e., on
15 Ni-NTA agarose (Qiagen, Valencia, CA) for the His,-tag and M2 Anti-Flag Agarose (Sigma, St. Louis, MO) for the FLAG-tag, following the procedures recommended by the manufacturers. The fragment-encoding genetic information recovered by KOH elution was amplified by PCR using Phannacia Ready-to-Go PCR Beads, 10 pmol of 5' and 3' PCR primers, and the following
20 PCR program (Pharmacia, Piscataway, NJ): Step 1: 95°C for 3 minutes; Step 2: 95°C for 30 seconds, 58/62°C for 30 seconds, 72°C for 1 minute, 20/25/30 cycles, as required; Step 3: 72°C for 5 minutes; Step 4: 4°C until end (typically, 25 cycles).

The resulting DNA was cleaved by 5-6 U EarI (New England Biolabs) per
25 µg DNA; the reaction took place in T4 DNA Ligase Buffer (New England Biolabs) at 37°C, for 1 hour, and was followed by an optional incubation at 70°C for 15 minutes to inactivate Ear I. Equal amounts of the BC, DE, and FG fragments were combined and ligated to form a full-length ¹⁰F_n3 gene with randomized loops. The ligation required 10 U of fresh EarI (New England

Biolabs) and 20 U of T4 DNA Ligase (Promega, Madison, WI), and took 1 hour at 37°C. EarI and ligase were then inactivated by a 15 minute incubation at 65°C.

Three different libraries, $BC_{wt}-DE_{wt}-FG_r$, $BC_r-DE_{wt}-FG_r$, and $BC_r-DE_r-FG_r$, were made in the manner described above. Each contained the form of the FG loop with 10 randomized residues. The BC and the DE loops of the first library bore the wild type $^{10}Fn3$ sequence; a BC loop with 7 randomized residues and a wild type DE loop made up the second library; and a BC loop with 7 randomized residues and a DE loop with 4 randomized residues made up the third library. The complexity of the FG loop in each of these three libraries was 10^{13} ; the further two randomized loops provided the potential for a complexity too large to be sampled in a laboratory. The combination of these libraries provided a master library having 10^{12} unique clones.

The sequences of 76 randomly picked clones from the original, randomized, $BC_r-DE_r-FG_r$ library showed no pattern in the randomized loops (data not shown); the amino acid frequency in the library varied in proportion to the number of codons available that encoded each residue, between 1% per position (glutamic acid, methionine, tryptophan) and ~~14%~~ per position (proline). In contrast, the average probability for a residue in the preserved, beta-sheet framework to have remained as wild type was found to be 99%.

Equimolar amounts of the three libraries (2 pmoles of DNA each) were combined into one master library in order to simplify the selection process; target binding itself was expected to select the most suitable library for a particular challenge. RNA-protein fusions were obtained from the master library following the general procedure described in Szostak et al., U.S.S.N. 09/007,005 and 09/247,190; Szostak et al., WO98/3 1700; and Roberts & Szostak, Proc. Natl. Acad. Sci. USA (1997) vol. 94, p. 12297-12302 (Figure S), except that affinity purification performed in rounds three to ten used only M2-Sepharose (see below).

Fusion Selections

The master library in the RNA-protein fusion form was subjected to selection for binding to TNF-a (Pepro Tech, Rocky Hill, NJ). Two initial protocols were employed: one in which the target was immobilized on an agarose column and one in which the target was immobilized on a BIACORE chip. First, an extensive optimization of conditions to minimize background binders to the agarose column yielded the favorable buffer conditions of 50 mM HEPES pH 7.4, 0.02% Triton, 100 µg/ml sheared salmon sperm DNA. In this buffer, the non-specific binding of the ¹⁰Fn3-RNA fusion to TNF-a Sepharose was 0.3%. The non-specific binding background of the ¹⁰Fn3-RNA/cDNA library to TNF-a Sepharose was found to be 0.1%.

During each round of selection on TNF-a Sepharose, the library was first preincubated for an hour with underivatized Sepharose to remove any remaining non-specific binders; the flow-through from this pre-clearing was incubated for another hour with TNF-a Sepharose. The TNF-a Sepharose was washed for 3-30 minutes.

After each selection, the cDNA component of the complex that had been eluted from the solid support with 0.3 M NaOH or 0.1M KOH was amplified by PCR, a DNA band of the expected size persisted through multiple rounds of selection (Figure 9); similar results were observed in the two alternative selection protocols, and only the data from the agarose column selection is shown in Figure 9.

In this selection, in the first seven rounds, the binding of Fn3-RNA/cDNA molecules to the target remained low; in contrast, when free protein was translated from DNA pools at different stages of the selection, the proportion of the column binding species increased significantly between rounds (Figure 10).

In later selections, the fusion pools selected in the first eight rounds of selection (R1-8) bound to TNF-a-Sepharose at levels close to the background (<0.25%) (Figure 13). After nine rounds of selection (R9), the binding of fusion to TNF-a-Sepharose increased sharply to 0.7%, and, after ten rounds of selection

(R10), the binding increased further to 7% (Figure 13). These selections were carried out using TNF- α immobilized on Epoxy-Activated Sepharose™ 6B (Amersham Pharmacia) at 10 mg TNF/g Sepharose in 10 mL. Before use, the TNF- α -derivatized Sepharose was blocked in Binding Buffer (50 mM HEPES, pH 7.4, 0.02% Triton, 0.1 mg/mL sheared salmon sperm DNA (Ambion)), overnight, at 4°C.

The ¹⁰Fn3-based master library was transcribed, ligated to the puromycin-bearing linker, translated into an mRNA-protein library in the presence of 5-10 μ L/300 μ L ³⁵S-methionine, affinity purified on Oligo(dT) Cellulose, reverse-transcribed into a DNA/mRNA-protein library, and affinity-purified on M2-Sepharose (for rounds 3-10), as described above. Forty pmol of DNA/mRNA-protein fusion library molecules, the equivalent of 20 copies of 4×10^{12} different sequences, were recovered, then subjected to the first round (R1) of the selection.

In the first step of the selection, 40 pmoles of the DNA/mRNA-protein library was incubated for 1 hour at 4°C, with tumbling, in 300 μ L of Binding Buffer with 30 μ L of Epoxy-Sepharose that had been subjected to the derivatization procedure in the absence of TNF- α . In the second round, 24 pmol of the library was added, and in the remaining eight rounds, 0.1-2 pmol of the library was added. The supernatant was recovered by microcentrifugation through a Micro Bio-Spin® chromatography column (BIO-RAD, Hercules, CA), then incubated with 30 μ L of TNF- α -Sepharose (6 μ M) in 300 μ L of the Binding Buffer for 1 hour at 4°C (during Rounds 7-10, the Binding Buffer contained an additional 1 mg/mL of BSA). The TNF- α -Sepharose was recovered on a spin column, then washed with 3 x 300 μ L of Binding Buffer, eluted with 100 μ L of 0.1 M KOH, and finally neutralized with 1 μ L of 1 M Tris 8.0, 8 μ L of 1 M HCl. Samples of the library, of the TNF- α -Sepharose before and after the elution, of the washes, and of the elutions were quantified by counting ³⁵S-methionine in the sample in a scintillation counter. The next round of selection began with the formation of a new DNA/mRNA-protein pool by

PCR amplification, which was transcribed, translated, and reverse-transcribed from the PCR product.

The DNA pools obtained from the elution after nine and after ten rounds were cloned into the TOPO™ TA®, pCR2.1 cloning vector (Invitrogen, Carlsbad, CA) and transformed into *E. coli*. Between 30 and 100 clones were picked and grown into plasmid minipreps (Qiagen). Thirty-eight clones from R9 and 29 clones from R10 were picked at random and sequenced (DNA Sequencing Core Facility, Massachusetts General Hospital, Dept. of Molecular Biology, Boston, MA). The program ClustalW⁶⁰ was used to align the resulting protein sequences.

Amino Acid Residue Sequences of the TNF- α Binding Clones

Thirty-eight of the 61 clones derived from R9 and from R10 had unique amino acid sequences, a surprising diversity. The ten clones that were isolated more than once, presumably because of their superior binding to TNF- α , are listed in Table 1 (full sequences in Figure 25).

Of the 61 clones picked randomly from the winning pool, only one (clone T09.08, sequence not shown) had its origin in the BC_{wt}-DE_{wt}-FG_r library, with another six from the BC_r-DE_{wt}-FG_r library. The observation that the remaining 54 (88% of the winners) were selected from the BC_r-DE_r-FG_r library points out the importance for TNF- α cooperative binding of the target by several loops.

The most common motif found in the selected loop sequences is PWA(S/T), which is found in the DE loop of 33 of the 61 clones; the more loosely defined sequence of PW(A/G) is seen in 41/61 clones. Such a strong selection for a specific DE sequence is surprising since the analogous CDR-H2 loops of antibody V_H domains generally make only a small contribution to antigen binding. On the other hand, the short length of the DE loop, which means that 10⁷ copies of each possible tetrapeptide sequence would be expected to be present in the library, would facilitate the optimization of any contribution of the DE loop to the selected properties. A survey of other Fn3 domains

(Dickenson et al., J. Mol. Biol. 236:1079-1092 (1994)) shows that proline is found at positions equivalent to the ¹⁰F_n3 residue 52 as frequently as is the wild-type glycine; similarly, alanine, glycine, and the wild-type lysine are all common at positions equivalent to the ¹⁰F_n3 position 54. In consequence, it appears likely that the selected residues at positions 52 and 54 are at least consistent with favorable biophysical properties. In contrast, no tryptophan is found at the position equivalent to the ¹⁰F_n3 residue 53, which suggests that Tryptophan 53 may have been selected for a reason specific to the present selection, such as due to a contribution to TNF-α binding. This is consistent with the absence of this motif in later selections against other antigens, again suggesting that the PWA/G motif is more likely to contribute to TNF-α binding directly than through stability or solubility of the ¹⁰F_n3 domain. The preference for the PWA/G motif on loop DE suggests another possible reason for the preference for the BC_r-DE_r-FG_r library during the selection: the BC_r-DE_r-FG_r library alone contained the randomized DE loop, and would be expected to outcompete the other two libraries if the PWA/G sequence were important to target binding.

The sequences selected most frequently in the BC loop is NRSGQLQS (12/61) (SEQ ID NO: 31), whereas the sequence selected most commonly in the FG loop is AQTGHHLHDK (6/61) (SEQ ID NO: 32). An NRSGQLQS BC loop and an AQTGHHLHDK FG loop have not been found in the same molecule, but two clones were found which contain the most frequently isolated sequences on two of the three randomized loops. These clones, T10.06 (BC: NRSGQLQS, DE: PWA) and T09.12 (DE: PWA, FG: AQTGHHLHDK), have two of the lowest four dissociation constants from TNF-α of the clones examined (Table 1).

Due to the use of a Taq polymerase that contains no proofreading activity, every round of PCR introduced additional random mutations into both the CDR-like loops and the beta-sheet scaffold of the ¹⁰F_n3 sequence, at the estimated rate of 0.01 % per base pair, i.e., 3% per ¹⁰F_n3 gene per round of PCR and approximately 75% per round of selection. Consequently, it is likely that

the residues preserved as wild-type and those preserved in a non-wild-type stable sequence indicate that such sequences were selected due to their superior properties. In the mutated loops, it is impossible to distinguish between the mutations introduced by oligonucleotide synthesis or by PCR mutagenesis, but
5 in the beta-strand scaffold, most of the mutations selected originate from Taq errors. The selected clones showed several conserved changes in the scaffold of the protein, which had not been randomized intentionally. Figure 18 indicates the residues in the ¹⁰F_n3 beta sheet that had not been randomized, but nevertheless mutated during selection. This mutagenesis occurred at the
10 frequency of 26-28 of the 61 clones; these mutations are marked with arrows under the wild-type ¹⁰F_n3 sequence and with the letter that identifies the selected residue. In particular, 28 of the 61 clones mutated from Leucine 18 to Arginine or to Glutamine, and 26 clones mutated from Threonine 56 to Isoleucine. Figure 19 shows the location of these scaffold mutations. Whereas position 56 is at the
15 stem loop DE and thus would be expected to affect the conformation and the target-binding properties of this loop, the distance of position 18 from the presumed TNF- α -binding loops suggests that the selective advantage of this mutation may arise from an indirect effect on the conformation of loop BC or from an effect on the stability of the protein (Figure 19). This is supported by an
20 experiment in which clone T10.06, which contains the frequently seen L18R and T56I changes from the wild-type, was mutagenized to reverse position 18 back to the wild-type leucine. This change caused an increase of the K_d of the variant by approximately 10-fold. The weaker binding of the T10.06(L18) protein to TNF- α suggests that the residue at position 18 has an effect on the binding of the
25 target by the CDR-like loops, possibly by a minor structural change that is transmitted through the beta-strand to loop BC.

Affinity and Specificity of the Selected TNF- α Binding Pools

The apparent average K_d values of free protein pools for TNF- α after nine and after ten rounds of selection were found to be indistinguishable (4 and 6 nM,

respectively; Table 1); this similarity in affinity is consistent with the relatively low (10 fold) level of enrichment observed in the last round of selection and with the similarity in the sequence composition of the two pools. The apparent average K_d values of free protein pool after four further rounds of selection was 5 3 nM, also indistinguishable from those of R9 and R10 pools

In order to assess the specificity of the binding of the pool selected after ten rounds of selection, we compared the binding of two different free protein pools to three cytokines immobilized on Sepharose to TNF- α , the target of the selection, and to IL-1 α and IL-13, which were irrelevant to the selection. The 10 first pool had been transcribed and translated from the initial, randomized DNA library before the selection (R0), and the second pool, from the library after ten rounds of selection (R10).

To carry out these experiments, the PCR product of the elution after the tenth round of selection was transcribed and translated *in vitro*, in the presence 15 of ^{35}S -methionine but without forming the mRNA-protein fusion. The resulting fraction of the free protein bound to TNF- α -Sepharose, to IL-1 α -Sepharose, to IL-13-Sepharose at approximately 10 μM , 30 μM , and 50 μM , respectively, and to underivatized Sepharose was compared (Figure 20), using the procedure described above for DNA/mRNA-protein fusion binding to TNF- α -Sepharose. 20 The amount of the selected pool bound to each of the targets was measured by scintillation counting of the washed beads.

Figure 20 shows that, whereas the binding of R0 to TNF- α , IL-1 α , and IL-13 was similar (2%, 4%, and 3%, respectively), the ten rounds of selection resulted in 32% binding to the targeted TNF- α , in 3% binding to IL-1 α , and in 25 1% binding to IL-13. The absolute and the relative increase of protein binding to TNF- α demonstrates the ability of the $^{10}\text{Fn3}$ scaffold and of the DNA/mRNA-protein fusion-based selection system to select target-specific binders.

To examine the specificity of binding further, clone T09.12 was 30 immobilized in a microarray format (as generally described below) and was

tested for binding to soluble TNF-a. Specific binding of TNF-a to this clone was detected using fluorescence detection (Figure 24A) and mass spectroscopy (Figure 24B). For the mass spectroscopy results, binding assays were carried out in the presence of fetal bovine serum, an exemplary complex biological fluid containing a variety of potential interfering proteins. For fluorescence detection (Figure 24A), a mixture of RNA-¹⁰F_n3 fusion of wild-type ¹⁰F_n3 and of the T09.12 variant (Table 1) was hybridized onto a DNA microarray on which oligonucleotides complementary to the RNA portion of the fusion molecules had been immobilized at 600 micron pitch, with 24 replicate features. After removal of unhybridized fusion by washing, the surface was exposed to biotin-TNF-a (2.6 µg/mL in TBS, 0.02% Tween-20, 0.2% BSA), washed, and air-dried. The captured biotin-TNF-a was detected by Cy3-labeled anti-biotin monoclonal antibody (Sigma) using a ScanArray 5000 system (GSI Lumonics). For mass spectroscopy detection, RNA-¹⁰F_n3 fusion of the T09.12 variant (Figure 24B) and wild-type ¹⁰F_n3 (Figure 24C) was treated with RNase A to generate a fusion between the protein and the DNA linker. The resulting DNA-linked protein was hybridized to a glass coverslip arrayed with an immobilized oligonucleotide complementary to the DNA linker (Figures 24B and 24C; no fusion was applied in Figure 24D). After washing, the above surfaces were exposed to TNF-a (1.5 mg/mL in 90% v/v PBS/10% fetal bovine serum). The dried chip was spotted with MALDI matrix and analyzed with a Voyager DE MALDI-TOF mass spectrometer (PerSeptive Biosystems). A signal at 17.4 kD, which corresponded to the expected molecular mass of monomeric TNF-a, was detected on the 200 µm features that contained T09.12 protein (Figure 24A), but not on the features that contained wild-type ¹⁰F_n3 (Figure 24B) nor on the features that did not contain DNA-protein fusion (Figure 24C).

K_d of the Selected TNF-a Binding Clones

Dissociation constants were determined for all the clones that were represented more than once in the two pools generated after nine and after ten

rounds of selection, as well as for the only clone that originated from the BC_{wt}-DE_{wt}-FG_{wt} library (clone T09.08).

To determine these binding constants, biotinylated TNF- α was prepared using the NHS-LC-LC-Biotin reagent supplied by Pierce (Rockford, IL).

- 5 MALDI-TOF mass spectrometry was used to estimate that more than 80% of the monomeric TNF- α , and hence more than 99% of the trimer, was biotinylated.

For the R9 and R10 pools (and the R14 and M12 pools discussed below), as well as for the characterized clones derived from these two pools, eleven samples of 0.25 nM, *in vitro*-translated, ³⁵S-methionine-labeled free protein were
10 incubated with the biotinylated TNF- α at a concentration between 17 pM and 23 nM, in 200 μ L 10 mM HEPES, pH 7.4, 150 mM NaCl, 1% BSA, 0.02% Triton, for one hour at room temperature. Subsequently, each sample was loaded on a pre-soaked, SAM^{2R} Biotin Capture Membrane (Promega, Madison, WI) using a 96 well, Easy-Titer™ ELIFA system (Pierce). Under vacuum, each spot was
15 washed with 200 μ L of HBS pH 7.4, 1% BSA, 0.05% Triton; next the entire membrane was rinsed in the buffer and air-dried. The membrane was exposed with a Storage Phosphor Screen (Molecular Dynamics, Sunnyvale, CA) overnight, and the intensities of the resulting individual spots were quantified using a STORM 860 phosphorimager with the ImageQuaNT densitometry
20 program (Molecular Dynamics). The K_d of the binding was determined by fitting the equilibrium equation to the resulting binding curve (KaleidaGraph, Synergy Software); the error of the experiment was estimated from 2-4 independent experiments.

In these studies, the K_d values were found to be in the narrow range of
25 1-24 nM (Table 1). The T09.12 and T10.06 clones, which contained the most commonly isolated sequences in two loops each, have the low K_d of 4 and 2 nM, respectively; on the other hand, a number of clones with less frequently seen loops, such as clones T09.07 and T10.15, showed similarly tight binding.

A sample comparison of TNF- α binding between free protein and the
30 cDNA/mRNA-protein complex derived from the same sequence showed that the

two dissociation constants were within experimental error of each other, a property of the system that makes it possible to use the cDNA/mRNA-protein complex to select for target-binding properties of the protein itself.

High-Stringency Selection of TNF- α Binding Clones

5 Despite the duplicate clones isolated, the TNF- α -binding pools after nine and after ten rounds of selection contained numerous different clones, i.e., 38 different sequences in 61 clones sampled. Therefore, further selection, with more stringent binding requirements, was undertaken to recover a subset of these clones with superior TNF- α binding properties. Consequently, four further
10 rounds of selections (R11-R14) were conducted in solution, where the concentration of the target was controlled more easily. The concentration of TNF- α was limited to 0.5 nM and the concentration of DNA/mRNA- $^{10}\text{Fn3}$ pool to 0.1 nM; in addition, the length and the temperature of the washes of the $^{10}\text{Fn3}$ /TNF- α complex bound to streptavidin-coated paramagnetic beads were
15 increased.

Specifically, these selections were carried out as follows. For rounds 11-13, 0.1 nM DNA/mRNA- $^{10}\text{Fn3}$ fusion library, which had been made as described above, was pre-cleared by tumbling for 1 hour at 4°C with 100 μL of Dynabeads® M-280 (streptavidin-coated paramagnetic beads; Dynal, Lake
20 Success, NY) that had been pre-blocked in Binding Buffer. The resulting pre-cleared fusion mixture was combined with 0.5 nM biotinylated TNF- α in 300 μL of the above Binding Buffer, and the complex incubated at 4°C for 1 hour. Next, 100 μL of resuspended Dynabeads® M-280 Streptavidin at 1.3 g/cm³, which had been blocked by overnight incubation in Binding Buffer, were added
25 to the mixture and incubated at 4°C, with tumbling, for 45 additional minutes. The paramagnetic beads were separated from the supernatant on a Dynal MPC-S rack, the supernatant was removed, and the beads were washed with the Binding Buffer for 1, 15, and 30 minutes in the case of R11 and R12, or for 1 minute, followed by nine ten-minute washes in the case of R13-R14. DNA was eluted

from the washed DNA/mRNA-¹⁰F_n3:TNF- α -biotin:streptavidin-bead complexes with two washes of 100 μ L 0.1 M KOH, and treated as described above for the column-based selection to produce the next generation DNA/mRNA-¹⁰F_n3 fusion library. Round 14 differed from R11-R13 in that the selection was
5 performed at 30°C and in the presence of an additional 150mM NaCl. Except for the elevated temperature, the sequence of washes was the same for R14 as for R13.

Twenty-two clones derived from the DNA eluted after four further rounds of selection (R14) were picked at random and found to represent 15 different
10 loop sequences (Table 2; full sequences in Figure 25). The clone T10.06, isolated previously from R10 as described above, was picked eight separate times, whereas the remaining sequences, including T09.31, which had been isolated from the R9 pool, were found in one isolate each. Similar to the isolates from rounds nine and ten, the R14 clones examined showed a preference (18 of
15 22 clones) for the PWA/G sequence in the DE loop, and four new, non-wild-type DE sequences were revealed.

Whereas the apparent average K_d values of the R14 free protein pool, 3 nM, is similar to those measured for the pools after nine and ten rounds (4 and 6 nM, respectively), several K_d values of the clones isolated from the R14 pool
20 were an order of magnitude lower than the lowest values observed in the R9 and R10 pools (Table 2). The clones that bound TNF- α most tightly, T14.07 and T14.25, had a K_d of 90 pmol. Thus, the conditions used in the last four rounds of selections were stringent enough to favor ¹⁰F_n3 molecules with subnanomolar K_d , but not so stringent as to eliminate such molecules.

25 Mutagenic Affinity Maturation

As discussed above, the selections described herein may also be combined with mutagenesis after all or a subset of the selection steps to further increase library diversity. In one parallel selection strategy, error-prone PCR was incorporated into the amplification of DNA between rounds (Cadwell and Joyce,

PCR Methods Appl 2:28 (1992)). This technique was carried out beginning with the diverse DNA pool eluted after R8 above. This pool was amplified using error-prone PCR, with the pool divided into seven equal parts and mutagenized at the target frequency of 0.8%, 1.6%, 2.4%, 3.2%, 4.0%, 4.8%,
5 and 5.6%. The seven PCR reactions were combined, and cDNA/RNA-protein fusion was made from the mixture and subjected to a round of selection in solution. Before the second mutagenic round, M10, error-prone PCR was performed in three separate reactions, at 0.8%, 1.6%, and 2.4%. The two remaining rounds, M11 and M12, were performed using standard Taq PCR.
10 Except for mutagenesis, the selection conditions for M9-M12 were the same as for R11-R14. The twenty M12 clones tested showed tighter binding to TNF- α than the clones selected using the two earlier selection protocols (Table 3; full sequences in Figure 25); the tightest binding of TNF- α was seen in M12.04, and had the observed K_d of 20 pM. These results demonstrated that low-level,
15 random mutagenesis late in a selection can improve both the binding affinity of selected antibody mimics (20 pM vs. 90 pM) and the speed with which they can be selected (12 rounds vs. 14 rounds). In addition, the frequency of tight binders in this mutagenesis approach was observed to be about 5%, whereas the frequency is approximately 3% in other selections.

20 Superiority of Fn Binders

The selection of $^{10}\text{Fn3}$ variants capable of binding to TNF- α , performed using covalent mRNA-protein fusion as the unit of selection, was won by molecules with dissociation constants as low as 20 pM. These K_d values compared favorably against the standards of selection of others that used other
25 antibody mimic scaffolds and selection methods. Consequently, the $^{10}\text{Fn3}$ -based scaffold and covalent mRNA-protein fusion-based *in vitro* selection method may be utilized for the development of antibody mimics against a broad range of antigens. In addition, the subnanomolar, TNF- α -binding $^{10}\text{Fn3}$ variants described herein represent potential therapeutic, research, and diagnostic agents.

Moreover, since this *in vitro* selection method can be automated, such a combination of scaffold and selection methods have applications on the genomic scale.

One of the factors that contributed to the success of the present selection was the randomization of all three CDR-like loops of ¹⁰F_n3; similar libraries which contained only one or two randomized loops were less likely to include tight binders than the library with three randomized, CDR-like loops.

In the selection reported above, the randomized loops remained the length of the corresponding, wild-type ¹⁰F_n3 loops. To generate further library diversity, the length of the loops as well as their sequences may be varied, to incorporate favorable mutations in the ¹⁰F_n3 beta-sheet into the wild-type scaffold used for library construction, and to create libraries with randomized beta-sheet scaffolds which will allow selection of structures even more successful at mimicking antibodies.

Selections similar to those described herein may be carried out with any other binding species target (for example, IL-1 or IL-13).

Animal Studies

Wild-type ¹⁰F_n3 contains an integrin-binding tripeptide motif, Arginine 78 - Glycine 79 - Aspartate 80 (the "RGD motif") at the tip of the FG loop. In order to avoid integrin binding and a potential inflammatory response based on this tripeptide *in vivo*, a mutant form of ¹⁰F_n3 was generated that contained an inert sequence, Serine 78 - Glycine 79 - Glutamate 80 (the "SGE mutant"), a sequence which is found in the closely related, wild-type ¹¹F_n3 domain. This SGE mutant was expressed as an N-terminally His₆-tagged, free protein in *E. coli*, and purified to homogeneity on a metal chelate column followed by a size exclusion column.

In particular, the DNA sequence encoding His₆-¹⁰F_n3(SGE) was cloned into the pET9a expression vector and transformed into BL21 DE3 pLysS cells. The culture was then grown in LB broth containing 50 µg/mL kanamycin at

37°C, with shaking, to $A_{560}=1.0$, and was then induced with 0.4 mM IPTG. The induced culture was further incubated, under the same conditions, overnight (14-18 hours); the bacteria were recovered by standard, low speed centrifugation. The cell pellet was resuspended in 1/50 of the original culture volume of lysis buffer (50 mM Tris 8.0, 0.5 M NaCl, 5% glycerol, 0.05% Triton X-100, and 1 mM PMSF), and the cells were lysed by passing the resulting paste through a Microfluidics Corporation Microfluidizer M1 10-EH, three times. The lysate was clarified by centrifugation, and the supernatant was filtered through a 0.45 μ m filter followed by filtration through a 0.2 μ m filter. 100 mL of the clarified lysate was loaded onto a 5 mL Talon cobalt column (Clontech, Palo Alto, CA), washed by 70 mL of lysis buffer, and eluted with a linear gradient of 0-30 mM imidazole in lysis buffer. The flow rate through the column through all the steps was 1 mL/min. The eluted protein was concentrated 10-fold by dialysis (MW cutoff = 3,500) against 15,000-20,000 PEG. The resulting sample was dialysed into buffer 1 (lysis buffer without the glycerol), then loaded, 5 mL at a time, onto a 16 x 60 mm Sephacryl 100 size exclusion column equilibrated in buffer 1. The column was run at 0.8 mL/min, in buffer 1; all fractions that contained a protein of the expected MW were pooled, concentrated 10X as described above, then dialyzed into PBS. Endotoxin screens and animal studies were performed on the resulting sample (Toxikon; MA).

The endotoxin levels in the samples examined to date have been below the detection level of the assay. In a preliminary animal toxicology study, this protein was injected into two mice at the estimated 100X therapeutic dose of 2.6 mg/mouse. The animals survived the two weeks of the study with no apparent ill effects. These safety results support the use of $^{10}\text{Fn3}$ incorporated into an IV drug.

Alternative Constructs for In Vivo Use

To extend the half life of the 8 kD $^{10}\text{Fn3}$ domain, a larger molecule has also been constructed that mimics natural antibodies. This $^{10}\text{Fn3-F}_c$ molecule

contains the $-\text{CH}_1-\text{CH}_2-\text{CH}_3$ (Figure 11) or $-\text{CH}_2-\text{CH}_3$ domains of the IgG constant region of the host; in these constructs, the $^{10}\text{Fn3}$ domain is grafted onto the N-terminus in place of the IgG V_H domain (Figures 11 and 12). Such antibody-like constructs are to improve the pharmacokinetics of the protein as well as its ability to harness the natural immune response.

In order to construct the murine form of the $^{10}\text{Fn3}-\text{CH}_1-\text{CH}_2-\text{CH}_3$ clone, the $-\text{CH}_1-\text{CH}_2-\text{CH}_3$ region was first amplified from a mouse liver spleen cDNA library (Clontech), then ligated into the pET25b vector. The primers used in the cloning were 5' Fc Nest and 3' 5 Fc Nest, and the primers used to graft the appropriate restriction sites onto the ends of the recovered insert were 5' Fc HIII and 3' Fc Nhe:

5' Fc Nest 5'GCG GCA GGG TTT GCT TAC TGG GGC CAA GGG 3' (SEQ ID NO: 15);

3' Fc Nest 5'GGG AGG GGT GGA GGT AGG TCA CAG TCC 3' (SEQ ID NO: 16);

3' Fc Nhe 5' TTT GCT AGC TTT ACC AGG AGA GTG GGA GGC 3' (SEQ ID NO: 17); and

5' Fc HIII 5' AAA AAG CTT GCC AAA ACG ACA CCC CCA TCT GTC 3' (SEQ ID NO: 18).

Further PCR was used to remove the CH_1 region from this clone and to create the Fc part of the shorter, $^{10}\text{Fn3}-\text{CH}_2-\text{CH}_3$ clone. The sequence encoding $^{10}\text{Fn3}$ was spliced onto the 5' end of each clone; either the wild type $^{10}\text{Fn3}$ cloned from the same mouse spleen cDNA library or a modified $^{10}\text{Fn3}$ obtained by mutagenesis or randomization of the molecules can be used. The oligonucleotides used in the cloning of murine wild-type $^{10}\text{Fn3}$ were:

Mo 5PCR-NdeI:

5' CATATGGTTTCTGATATTCGAGAGATCTGGAG 3' (SEQ ID NO: 19);

Mo5PCR-His-NdeI (for an alternative N-terminus with the His, purification tag):

5' CAT ATG CAT CAC CAT CAC CAT CAC GTT TCT GAT ATT CCG AGA G 3' (SEQ ID NO: 20); and

5 Mo3PCR-EcoRI: 5' GAATTCCTATGTTTTATAATTGATGGAAAC3' (SEQ ID NO: 21).

The human equivalents of the clones are constructed using the same strategy with human oligonucleotide sequences.

Antibody Mimics in Protein Chip Applications

10 Any of the antibody mimics described herein may be immobilized on a solid support, such as a microchip. The suitability of the present scaffolds, for example, the ¹⁰F_n3 scaffold, for protein chip applications is the consequence of (1) their ability to support many binding functions which can be selected rapidly on the bench or in an automated setup, and (2) their superior biophysical
15 properties.

The versatile binding properties of ¹⁰F_n3 are a function of the loops displayed by the F_n3 immunoglobulin-like, beta sandwich fold. As discussed above, these loops are similar to the complementarity determining regions of antibody variable domains and can cooperate in a way similar to those antibody
20 loops in order to bind antigens. In our system, ¹⁰F_n3 loops BC (for example, residues 21-30), DE (for example, residues 51-56), and FG (for example, residues **76-87**) are randomized either in sequence, in length, or in both sequence and length in order to generate diverse libraries of mRNA-¹⁰F_n3 fusions. The binders in such libraries are then enriched based on their affinity
25 for an immobilized or tagged target, until a small population of high affinity binders are generated. Also, error-prone PCR and recombination can be employed to facilitate affinity maturation of selected binders. Due to the rapid and efficient selection and affinity maturation protocols, binders to a large number of targets can be selected in a short time.

As a scaffold for binders to be immobilized on protein chips, the ¹⁰F_n3 domain has the advantage over antibody fragments and single-chain antibodies of being smaller and easier to handle. For example, unlike single-chain scaffolds or isolated variable domains of antibodies, which vary widely in their stability and solubility, and which require an oxidizing environment to preserve their structurally essential disulfide bonds, ¹⁰F_n3 is extremely stable, with a melting temperature of 110°C, and solubility at a concentration > 16 mg/mL. The ¹⁰F_n3 scaffold also contains no disulfides or free cysteines; consequently, it is insensitive to the redox potential of its environment. A further advantage of ¹⁰F_n3 is that its antigen-binding loops and N-terminus are on the edge of the beta-sandwich opposite to the C-terminus; thus the attachment of a ¹⁰F_n3 scaffold to a chip by its C-terminus aligns the antigen-binding loops, allowing for their greatest accessibility to the solution being assayed. Since ¹⁰F_n3 is a single domain of only 94 amino acid residues, it is also possible to immobilize it onto a chip surface at a higher density than is used for single-chain antibodies, with their approximately 250 residues. In addition, the hydrophilicity of the ¹⁰F_n3 scaffold, which is reflected in the high solubility of this domain, minimizes unwanted binding of ¹⁰F_n3 to a chip surface.

The stability of the ¹⁰F_n3 scaffold as well as its suitability for library formation and selection of binders are likely to be shared by the large, F_n3-like class of protein domains with an immunoglobulin-like fold, such as the domains of tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-R, cytokine receptor, glycosidase inhibitor, and antibiotic chromoprotein. The key features shared by all such domains are a stable framework provided by two beta-sheets, which are packed against each other and which are connected by at least three solvent-accessible loops per edge of the sheet; such loops can be randomized to generate a library of potential binders without disrupting the structure of the framework (as described above). In addition, as with ¹⁰F_n3, any of these loops (or similar loops from other proteins) may be immobilized alone or in combination with other loops onto a solid support surface.

Immobilization of Fn3-Based Antibody Mimics

- To immobilize antibody mimics, such as Fn3-based antibody mimics, to a chip surface, a number of exemplary techniques may be utilized. For example, such antibody mimics may be immobilized as RNA-protein fusions by
- 5 Watson-Crick hybridization of the RNA moiety of the fusion to a base complementary DNA immobilized on the chip surface (as described, for example, in Addressable Protein Arrays, U.S.S.N. 60/080,686; U.S.S.N. 09/282,734; and WO 99/5 1773; and Methods for Encoding and Sorting In Vitro Translated Proteins, U.S.S.N. 60/15 1,261 and U.S.S.N. 09/648,040).
- 10 Alternatively, antibody mimics can be immobilized as free proteins directly on a chip surface. Manual as well as robotic devices may be used for deposition of the antibody mimics on the chip surface. Spotting robots can be used for deposition of antibody mimics with high density in an array format (for example, by the method of Lueking et al., Anal Biochem. 1999 May
- 15 15;270(1):103-11). Different methods may also be utilized for anchoring the antibody mimic on the chip surface. A number of standard immobilization procedures may be used including those described in Methods in Enzymology (K. Mosbach and B. Danielsson, eds.), vols. 135 and 136, Academic Press, Orlando, Florida, 1987; Nilsson et al., Protein Expr. Purif. 1997 Oct; 11(1): 1-16;
- 20 and references therein. Oriented immobilization of antibody mimics can help to increase the binding capacity of chip-bound antibody mimics. Exemplary approaches for achieving oriented coupling are described in Lu et al., The Analyst (1996), vol. 121, p. 29R-32R; and Turkova, J Chromatogr B Biomed Sci App. 1999 Feb 5;722(1-2):11-31. In addition, any of the methods described
- 25 herein for anchoring antibody mimics to chip surfaces can also be applied to the immobilization of antibody mimics on beads, or other supports.

Target Protein Capture and Detection

Selected populations of scaffold-binders may be used for detection and/or quantitation of analyte targets, for example, in samples such as biological

samples. To carry out this type of diagnostic assay, selected scaffold-binders to targets of interest are immobilized on an appropriate support to form multi-featured protein chips. Next, a sample is applied to the chip, and the components of the sample that associate with the binders are identified based on the target-specificity of the immobilized binders. Using this technique, one or more components may be simultaneously identified or quantitated in a sample (for example, as a means to carry out sample profiling).

Methods for target detection allow measuring the levels of bound protein targets and include, without limitation, radiography, fluorescence scanning, mass spectroscopy (MS), and surface plasmon resonance (SPR).

Autoradiography using a phosphorimager system (Molecular Dynamics, Sunnyvale, CA) can be used for detection and quantification of target protein which has been radioactively labeled, e.g., using ^{35}S methionine. Fluorescence scanning using a laser scanner (see below) may be used for detection and quantification of fluorescently labeled targets. Alternatively, fluorescence scanning may be used for the detection of fluorescently labeled ligands which themselves bind to the target protein (e.g., fluorescently labeled target-specific antibodies or fluorescently labeled streptavidin binding to target-biotin, as described below).

Mass spectroscopy can be used to detect and identify bound targets based on their molecular mass. Desorption of bound target protein can be achieved with laser assistance directly from the chip surface as described below. Mass detection also allows determinations, based on molecular mass, of target modifications including post-translational modifications like phosphorylation or glycosylation. Surface plasmon resonance can be used for quantification of bound protein targets where the scaffold-binder(s) are immobilized on a suitable gold-surface (for example, as obtained from Biacore, Sweden).

Described below are exemplary schemes for selecting binders (in this case, Fn-binders specific for the protein, TNF- α) and the use of those selected

populations for detection on chips. This example is provided for the purpose of illustrating the invention, and should not be construed as limiting.

Selection of TNF- α Binders Based on 10 F α 3 Scaffold

In one exemplary use for scaffold selection on chips, an 10 F α 3-based
5 selection was performed against TNF- α , using a library of human 10 F α 3 variants with randomized loops BC, DE, and FG. The library was constructed from three DNA fragments, each of which contained nucleotide sequences that encoded approximately one third of human 10 F α 3, including one of the randomized loops. The DNA sequences that encoded the loop residues listed above were rebuilt by
10 oligonucleotide synthesis, so that the codons for the residues of interest were replaced by (NNS) $_n$, where N represents any of the four deoxyribonucleotides (A, C, G, or T), and S represents either C or G. The C-terminus of each fragment contained the sequence for the FLAG purification tag.

Once extended by Klenow, each DNA fragment was transcribed, ligated to
15 a puromycin-containing DNA linker, and translated *in vitro*, as described by Szostak et al. (Roberts and Szostak, Proc. Natl. Acad. Sci USA 94:12297, 1997; Szostak et al., U.S.S.N. 09/007,005 and U.S.S.N. 09/247,190; Szostak et al., WO98/31700), to generate an mRNA-peptide fusion, which was then reverse-transcribed into a DNA-mRNA-peptide fusion. The binding of the
20 FLAG-tagged peptide to M2 agarose separated full-length fusion molecules from those containing frameshifts or superfluous stop codons; the DNA associated with the purified full-length fusion was amplified by PCR, then the three DNA fragments were cut by Ear I restriction endonuclease and ligated to form the full length template. The template was transcribed, ligated to
25 puromycin-containing DNA linkers, and translated to generate a 10 F α 3-RNA/cDNA library, which was then reverse-transcribed to yield the DNA-mRNA-peptide fusion library which was subsequently used in the selection.

Selection for TNF- α binders took place in 50 mM HEPES, pH 7.4, 0.02% Triton-X, 0.1 mg/mL salmon sperm DNA. The PROfusion™ library was incubated with Sepharose-immobilized TNF- α ; after washing, the DNA associated with the tightest binders was eluted with 0.1 M KOH, amplified by
5 PCR, and transcribed, ligated, translated, and reverse-transcribed into the starting material for the next round of selection.

Ten rounds of such selection were performed (as shown in Figure 13); they resulted in a PROfusion™ pool that bound to TNF- α -Sepharose with the apparent average K_d of 120 nM. Specific clonal components of the pool that
10 were characterized showed TNF- α binding in the range of 50-500 nM.

Immobilization, Target Protein Capture, and MALDI-TOF Detection

As a first step toward immobilizing Fn3 fusions to a chip surface, an oligonucleotide capture probe was prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using the solid-support phosphoramidite
15 approach. All reagents were obtained from Glen Research. Synthesis was initiated with a solid support containing a disulfide bond to eventually provide a 3'-terminal thiol functionality. The first four monomers to be added were hexaethylene oxide units, followed by 20 T monomers. The 5'-terminal DMT group was not removed. The capture probe was cleaved from the solid support
20 and deprotected with ammonium hydroxide, concentrated to dryness in a vacuum centrifuge, and purified by reverse-phase HPLC using an acetonitrile gradient in triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and the 5'-terminal DMT group was removed by treatment with 80% AcOH for 30
25 minutes. The acid was removed by evaporation, and the oligonucleotide was then treated with 100 mM DTT for 30 minutes to cleave the disulfide bond. DTT was removed by repeated extraction with EtOAc. The oligonucleotide was ethanol precipitated from the remaining aqueous layer and checked for purity by reverse-phase HPLC.

- The 3'-thiol capture probe was adjusted to 250 μ M in degassed 1X PBS buffer and applied as a single droplet (75 μ L) to a 9x9mm gold-coated chip (Biacore) in an argon-flushed chamber containing a small amount of water. After 18 hours at room temperature, the capture probe solution was removed,
- 5 and the functionalized chip was washed with 50 mL 1X PBS buffer (2x for 15 minutes each) with gentle agitation, and then rinsed with 50 mL water (2x for 15 minutes each) in the same fashion. Remaining liquid was carefully removed and the functionalized chips were either used immediately or stored at 4°C under argon.
- 10 About 1pmol of 10 F_n3 fusion pool from the Round 10 TNF- α selection (above) was treated with RNase A for several hours, adjusted to 5X SSC in 70 μ L, and applied to a functionalized gold chip from above as a single droplet. A 50 μ L volume gasket device was used to seal the fusion mixture with the functionalized chip, and the apparatus was continuously rotated at 4°C. After
- 15 18 hours the apparatus was disassembled, and the gold chip was washed with 50 mL 5X SSC for 10 minutes with gentle agitation. Excess liquid was carefully removed from the chip surface, and the chip was passivated with a blocking solution (1X TBS + 0.02% Tween-20 + 0.25% BSA) for 10 minutes at 4°C. Excess liquid was carefully removed, and a solution containing 500 μ g/mL
- 20 TNF- α in the same composition blocking solution was applied to the chip as a single droplet and incubated at 4°C for two hours with occasional mixing of the droplet via Pipetman. After removal of the binding solution, the chip was washed for 5 minutes at 4°C with gentle agitation (50 mL 1X TBS + 0.02% Tween-20) and then dried at room temperature. A second chip was prepared
- 25 exactly as described above, except fusion was not added to the hybridization mix.

Next, MALDI-TOF matrix (15 mg/mL 3,5-dimethoxy-4-hydroxycinnamic acid in 1:1 ethanol/10% formic acid in water) was uniformly applied to the gold chips with a high-precision 3-axis robot (MicroGrid, BioRobotics). A 16-pin

30 tool was used to transfer the matrix from a 384-well microtiter plate to the chips,

producing 200 micron diameter features with a 600 micron pitch. The MALDI-TOF mass spectrometer (Voyager DE, PerSeptive Biosystems) instrument settings were as follows: Accelerating Voltage = 25k, Grid Voltage = 92%, Guide Wire Voltage = 0.05%, Delay = 200 on, Laser Power = 2400, 5 Low Mass Gate = 1500, Negative Ions = off. The gold chips were individually placed on a MALDI sample stage modified to keep the level of the chip the same as the level of the stage, thus allowing proper flight distance. The instrument's video monitor and motion control system were used to direct the laser beam to individual matrix features.

10 Figures 14 and 15 show the mass spectra from the $^{10}\text{Fn}3$ fusion chip and the non-fusion chip, respectively. In each case, a small number of 200 micron features were analyzed to collect the spectra, but Figure 15 required significantly more acquisitions. The signal at 17.4 kDa corresponds to TNF- α monomer.

Immobilization, Target Protein Capture, and Fluorescence Detection

15 Pre-cleaned 1x3 inch glass microscope slides (Goldseal, #3010) were treated with Nanostrip (Cyantek) for 15 minutes, 10% aqueous NaOH at 70°C for 3 minutes, and 1% aqueous HCl for 1 minute, thoroughly rinsing with deionized water after each reagent. The slides were then dried in a vacuum desiccator over anhydrous calcium sulfate for several hours. A 1% solution of 20 aminopropyltrimethoxysilane in 95% acetone / 5% water was prepared and allowed to hydrolyze for 20 minutes. The glass slides were immersed in the hydrolyzed silane solution for 5 minutes with gentle agitation. Excess silane was removed by subjecting the slides to ten 5-minute washes, using fresh portions of 95% acetone / 5% water for each wash, with gentle agitation. The 25 slides were then cured by heating at 110°C for 20 minutes. The silane treated slides were immersed in a freshly prepared 0.2% solution of phenylene 1,4-diisothiocyanate in 90% DMF / 10% pyridine for two hours, with gentle agitation. The slides were washed sequentially with 90% DMF / 10% pyridine, methanol, and acetone. After air drying, the functionalized slides were stored at

0°C in a vacuum desiccator over anhydrous calcium sulfate. Similar results were obtained with commercial amine-reactive slides (3-D Link, Surmodics).

Oligonucleotide capture probes were prepared with an automated DNA synthesizer (PE BioSystems Expedite 8909) using conventional phosphoramidite chemistry. All reagents were from Glen Research. Synthesis was initiated with a solid support bearing an orthogonally protected amino functionality, whereby the 3'-terminal amine is not unmasked until final deprotection step. The first four monomers to be added were hexaethylene oxide units, followed by the standard A, G, C and T monomers. All capture
10 oligo sequences were cleaved from the solid support and deprotected with ammonium hydroxide, concentrated to dryness, precipitated in ethanol, and purified by reverse-phase HPLC using an acetonitrile gradient in triethylammonium acetate buffer. Appropriate fractions from the HPLC were collected, evaporated to dryness in a vacuum centrifuge, and then coevaporated
15 with a portion of water.

The purified, amine-labeled capture oligos were adjusted to a concentration of 250 μ M in 50 mM sodium carbonate buffer (pH 9.0) containing 10% glycerol. The probes were spotted onto the amine-reactive glass surface at defined positions in a 5x5x6 array pattern with a 3-axis robot (MicroGrid,
20 BioRobotics). A 16-pin tool was used to transfer the liquid from 384-well microtiter plates, producing 200 micron features with a 600 micron pitch. Each sub-grid of 24 features represents a single capture probe (i.e., 24 duplicate spots). The arrays were incubated at room temperature in a moisture-saturated environment for 12-18 hours. The attachment reaction was terminated by
25 immersing the chips in 2% aqueous ammonium hydroxide for five minutes with gentle agitation, followed by rinsing with distilled water (3X for 5 minutes each). The array was finally soaked in 10X PBS solution for 30 minutes at room temperature, and then rinsed again for 5 minutes in distilled water.

Specific and thermodynamically isoenergetic sequences along the ¹⁰F_n3 mRNA were identified to serve as capture points to self-assemble and anchor the ¹⁰F_n3 protein. The software program HybSimulator v4.0 (Advanced Gene Computing Technology, Inc.) facilitated the identification and analysis of

5 potential capture probes. Six unique capture probes were chosen and printed onto the chip, three of which are complementary to common regions of the ¹⁰F_n3 fusion pool's mRNA (CP3', CP5', and CPflag). The remaining three sequences (CPneg1, CPneg2, and CPneg3) are not complementary and function in part as negative controls. Each of the capture probes possesses a 3'-amino terminus and

10 four hexaethylene oxide spacer units, as described above. The following is a list of the capture probe sequences that were employed (5'-3'):

CP3': TGTAATAGTAATTGTCCC (SEQ ID NO: 22)
 CP5': TTTTTTTTTTTTTTTTTTTT (SEQ ID NO: 23)
 CPneg1: CCTGTAGGTGTCCAT (SEQ ID NO: 24)
 15 CPflag: CATCGTCCTTGTAGTC (SEQ ID NO: 25)
 CPneg2: CGTCGTAGGGGTA (SEQ ID NO: 26)
 CPneg3: CAGGTCTTCTTCAGAGA (SEQ ID NO: 27)

About 1pmol of ¹⁰F_n3 fusion pool from the Round 10 TNF-α selection was adjusted to 5X SSC containing 0.02% Tween-20 and 2 mM vanadyl

20 ribonucleotide complex in a total volume of 350 μL. The entire volume was applied to the microarray under a 400 μL gasket device and the assembly was continuously rotated for 18 hours at room temperature. After hybridization the slide was washed sequentially with stirred 500 mL portions of 5X SSC, 2.5X SSC, and 1X SSC for 5 minutes each. Traces of liquid were removed by

25 centrifugation and the slide was allowed to air-dry.

Recombinant human TNF-α (500 μg, lyophilized, from PreproTech) was taken up in 230 μL 1X PBS and dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The dialyzed TNF-

a was treated with EZ-Link NHS-LC-LC biotinylation reagent (20 µg, Pierce) for 2 hours at 0°C, and again dialyzed against 700 mL stirred 1X PBS at 4°C for 18 hours in a Microdialyzer unit (3,500 MWCO, Pierce). The resulting conjugate was analyzed by MALDI-TOF mass spectrometry and was found to
5 be almost completely functionalized with a single biotin moiety.

Each of the following processes was conducted at 4°C with continuous rotation or mixing. The protein microarray surface was passivated by treatment with 1X TBS containing 0.02% Tween-20 and 0.2% **BSA** (200 µL) for 60 minutes. Biotinylated TNF-α (100 nM concentration made **up** in the passivation
10 buffer) was contacted with the microarray for 120 minutes. The microarray was washed with 1X TBS containing 0.02% Tween-20 (3X 50 mL, 5 minutes each wash). Fluorescently labeled streptavidin (2.5 µg/mL Alexa 546-streptavidin conjugate from Molecular Probes, made up in the passivation buffer) was contacted with the microarray for 60 minutes. The microarray was washed with
15 1X TBS containing 0.02% Tween-20 (2X 50 mL, 5 minutes each wash) followed by a 3 minute rinse with 1X TBS. Traces of liquid were removed by centrifugation, and the slide was allowed to air-dry at room temperature.

Fluorescence laser scanning was performed with a GSI Lumonics ScanArray 5000 system using 10 µM pixel resolution and preset excitation and
20 emission wavelengths for Alexa 546 dye. Phosphorimage analysis was performed with a Molecular Dynamics Storm system. Exposure time was 48 hours with direct contact between the microarray and the phosphor storage screen. Phosphorimage scanning was performed at the 50 micron resolution setting, and data was extracted with ImageQuant v.4.3 software.

25 Figures 16 and 17 are the phosphorimage and fluorescence scan, respectively, of the same array. The phosphorimage shows where the ¹⁰F_n3 fusion hybridized based on the ³⁵S methionine signal. The fluorescence scan shows where the labeled TNF-α bound.

Other Embodiments

Other embodiments are within the claims.

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference.

5 What is claimed is:

Claims

1. A non-antibody derivative protein comprising a domain having an immunoglobulin-like fold, said non-antibody protein deriving from a reference protein by having a mutated amino acid sequence wherein said non-antibody
- 5 protein binds with a K_d at least as tight as 10 nM to a compound that is not bound as tightly by said reference protein.
2. The derivative protein of claim 1, said derivative protein binding with K_d at least as tight as 1 nM, at least as tight as 500 pM, at least as tight as 100 pM, or at least as tight as 20 pM.
- 10 3. The derivative protein of claim 1, wherein said derivative protein contains one, two, or three mutated loops and wherein at least one of said loops contributes to the binding of said derivative protein to said compound.
4. The derivative protein of claim 3, wherein at least two of said mutated loops or three of said mutated loops contribute to said binding of said derivative
- 15 protein to said compound.
5. The derivative protein of claim 1, wherein said reference protein lacks disulfide bonds.
6. The derivative protein of claim 1, wherein said derivative protein has at least one disulfide bond.
- 20 7. The derivative protein of claim 1, wherein said domain having an immunoglobulin-like fold has a molecular mass less than 10 kD, greater than 7.5 kD, or between 7.5-10kD.

8. The derivative protein of claim 1, wherein said derivative protein is a monomer or dimer under physiological conditions.
9. The derivative protein of claim 1, wherein said reference protein is a naturally-occurring mammalian protein.
- 5 10. The derivative protein of claim 1, wherein said domain having an immunoglobulin-like fold is mutated and comprises up to 34% mutated amino acids as compared to the immunoglobulin-like fold of said reference protein.
11. The derivative protein of claim 1, wherein said domain having an immunoglobulin-like fold consists of approximately 50-150 amino acids, and
10 preferably approximately 50 amino acids.
12. The derivative protein of claim 1, said derivative protein being immobilized on a solid support.
13. The derivative protein of claim 12, said derivative protein being part of an array immobilized on said solid support.
- 15 14. The derivative protein of claim 12, said solid support being a chip or bead.
15. The derivative protein of claim 1, wherein said derivative protein is joined to a heterologous protein, said heterologous protein being unable to bind to said compound.

16. The derivative protein of claim 15, wherein said heterologous protein is an antibody or an antibody domain, comprises an immunoglobulin F_c domain, or is selected from the group consisting of a complement protein, a toxin protein, and an albumin protein.
- 5 17. The derivative protein of claim 1, wherein said derivative protein is covalently bound to a nucleic acid.
18. The derivative protein of claim 17, wherein said nucleic acid encodes said protein.
19. The derivative protein of claim 1, wherein said derivative protein is
10 derived from a fibronectin or fibronectin dimer, tenascin, N-cadherin, E-cadherin, ICAM, titin, GCSF-receptor, cytokine receptor, glycosidase inhibitor, antibiotic chromoprotein, myelin membrane adhesion molecule PO, CD8, CD4, CD2, class I MHC, T-cell antigen receptor, CD1, C2 and I-set domains of VCAM-1, I-set immunoglobulin domain of myosin-binding protein
15 C, I-set immunoglobulin domain of myosin-binding protein H, I-set immunoglobulin domain of telokin, NCAM, twitchin, neuroglian, growth hormone receptor, erythropoietin receptor, prolactin receptor, interferon-gamma receptor, β -galactosidase/glucuronidase, β -glucuronidase, transglutaminase, T-cell antigen receptor, superoxide dismutase, tissue factor domain, cytochrome F,
20 green fluorescent protein, GroEL, or thaumatin.
20. A nucleic acid encoding a derivative protein of claim 1.

21. A method of obtaining a derivative non-antibody protein which binds to a compound, said method comprising:
- (a) providing a non-antibody scaffold protein comprising an immunoglobulin-like fold, wherein said scaffold protein does not bind to said
 - 5 compound with a K_d as tight as 10 nM;
 - (b) generating mutated derivatives of said non-antibody scaffold protein, thereby producing a library of mutated proteins;
 - (c) contacting said library with said compound;
 - (d) selecting from said library at least one derivative protein which binds
 - 10 to said compound with a K_d at least as tight as 10 nM; and
 - (e) optionally repeating steps (b) - (d) substituting for the non-antibody scaffold protein in repeated step (b) the product from the previous step (d).

22. A method for obtaining a non-antibody protein which binds to a compound, said method comprising:
- 15 (a) contacting said compound with a candidate protein, said candidate protein being a derivative non-antibody protein comprising a domain having an immunoglobulin-like fold, said non-antibody protein deriving from a reference protein by having a mutated amino acid sequence wherein said non-antibody protein binds with a K_d at least as tight as 10 nM to a compound that is not
 - 20 bound as tightly by said reference protein, wherein said contacting is carried out under conditions that allow compound-protein complex formation; and
 - (b) obtaining, from said complex, said derivative protein which binds to said compound.

23. A method for obtaining a compound which binds to a non-antibody protein, said non-antibody protein comprising a domain having an immunoglobulin-like fold and deriving from a reference protein by having a mutated amino acid sequence, wherein said non-antibody protein binds with a
5 Kd at least as tight as 10nM to a compound that is not bound as tightly by said reference protein, said method comprising:

- (a) contacting said derivative protein with a candidate compound, wherein said contacting is carried out under conditions that allow compound-protein complex formation; and
- 10 (b) obtaining, from said complex, said compound which binds to said derivative protein.

24. A method for detecting a compound in a sample, said method comprising:

- (a) contacting said sample with a non-antibody protein comprising a
15 domain having an immunoglobulin-like fold, said non-antibody protein deriving from a reference protein by having a mutated amino acid sequence wherein said non-antibody protein binds with a Kd at least as tight as 10nM to a compound that is not bound as tightly by said reference protein, wherein said contacting is carried out under conditions that allow compound-protein complex formation;
- 20 and
- (b) detecting said complex, thereby detecting said compound in said sample.

25. A non-antibody protein that binds tumor necrosis factor- α (TNF- α) with a Kd at least as tight as 1 μ M, said protein having a sequence that is less
25 than 20% identical to TNF- α receptor.

26. The non-antibody protein of claim 25, wherein said non-antibody protein comprises a mutated fibronectin type III domain and wherein said protein is mutated in the DE, BC, and FG loops.

27. The non-antibody protein of claim 26, wherein said FG loop is the
5 same length as the wild-type FG loop.

28. The non-antibody protein of claim 25, wherein said non-antibody protein comprises an immunoglobulin-like fold.

29. The non-antibody protein of claim 28, wherein said immunoglobulin-like fold has a molecular mass less than 10kD, greater than 7.5 kD, or between
10 7.5-10 kD.

30. The non-antibody protein of claim 29, wherein said immunoglobulin-like fold consists of approximately 50-150 amino acids, and preferably, approximately 50 amino acids.

15 31. The non-antibody protein of claim 25, wherein said non-antibody protein binds TNF- α with a K_d at least as tight as 500 nM, 100 nM, 10 nM, 1 nM, 500 pM, 100 pM, or 20 pM.

32. The non-antibody protein of claim 25, wherein said non-antibody protein contains one, two, or three mutated loops.

20 33. The non-antibody protein of claim 32, wherein at least one, and preferably two or all three of said mutated loops, contribute to the binding of the non-antibody protein to TNF- α .

34. The non-antibody protein of claim 25, wherein said non-antibody protein has at least one disulfide bond.

35. The non-antibody protein of claim 25, wherein said non-antibody protein is a monomer or dimer under physiological conditions.

5 36. The non-antibody protein of claim 25, wherein said non-antibody protein is immobilized on a solid support.

37. The non-antibody protein of claim 36, wherein said solid support is a chip or bead.

38. The non-antibody protein of claim 36, wherein said non-antibody
10 protein is part of an array immobilized on said solid support.

39. The non-antibody protein of claim 25, wherein said non-antibody protein is joined to a heterologous protein.

40. The non-antibody protein of claim 39, wherein said heterologous protein is an antibody or an antibody domain that does not bind TNF- α , an
15 immunoglobulin F_c domain, a complement protein, or an albumin protein.

41. The non-antibody protein of claim 25, wherein said non-antibody protein comprises a mutated fibronectin type III domain and preferably a mutated human fibronectin type III domain.

42. The non-antibody protein of claim 41, wherein said mutated
20 fibronectin type III domain is a mutated tenth module of the fibronectin type III domain (¹⁰F_n3))

43. The non-antibody protein of claim 42, wherein said non-antibody protein lacks an ¹⁰Fn3 integrin-binding motif.

44. The non-antibody protein of claim 42, wherein said non-antibody protein comprises a non-naturally occurring sequence in a loop of ¹⁰Fn3, and
5 preferably, the loop sequence PW(A/G), and/or comprises a non-naturally occurring sequence in a β -sheet of ¹⁰Fn3.

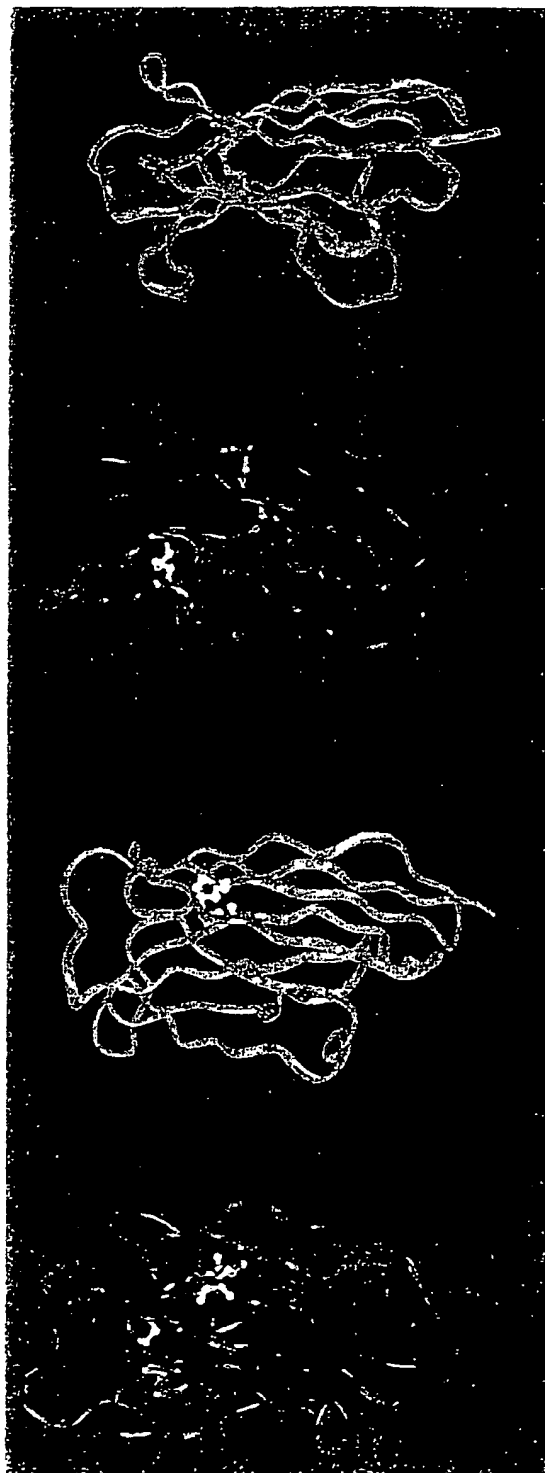
45. The non-antibody protein of claim 25, wherein said non-antibody protein comprises any one of the sequences of Figure 25 (SEQ ID NOS: 34-140).

10 46. A nucleic acid encoding a non-antibody protein of claim 25.

47. A loop structure on a protein, said loop comprising any one of the amino acid sequences of Figure 25 (SEQ ID NOS: 34-140).

1/25

FIG. 1



2/25

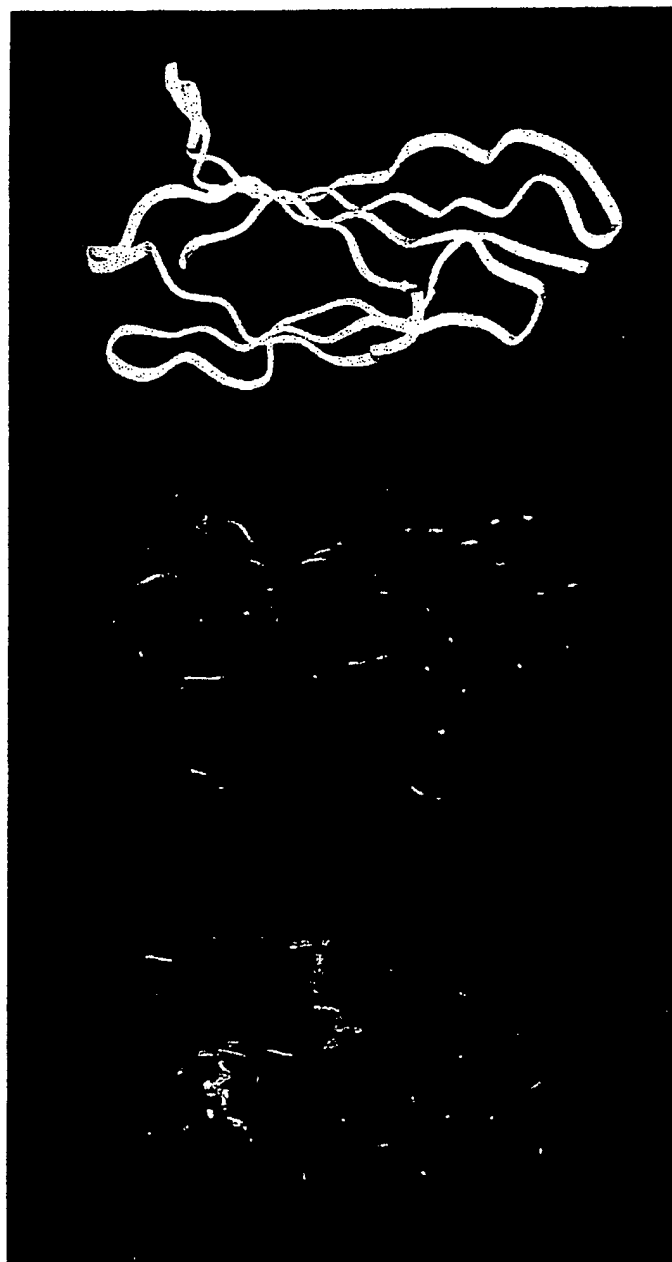
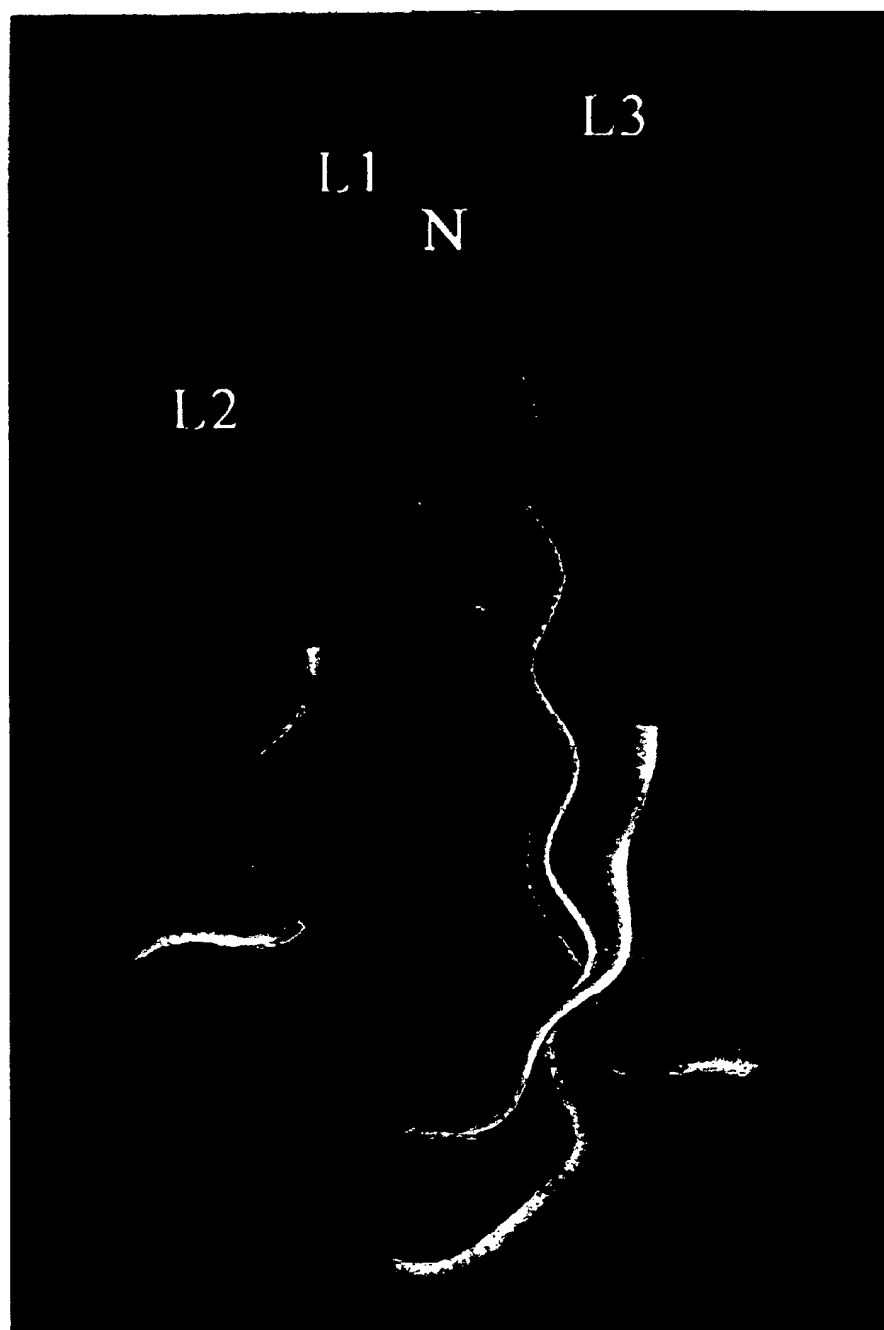


FIG. 2

3/25

FIG. 3



5/25

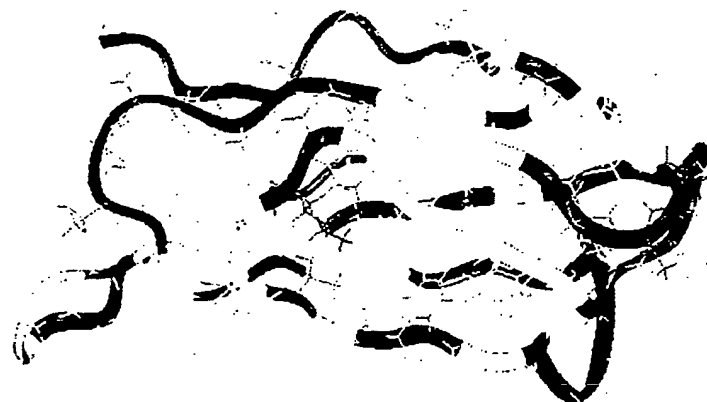
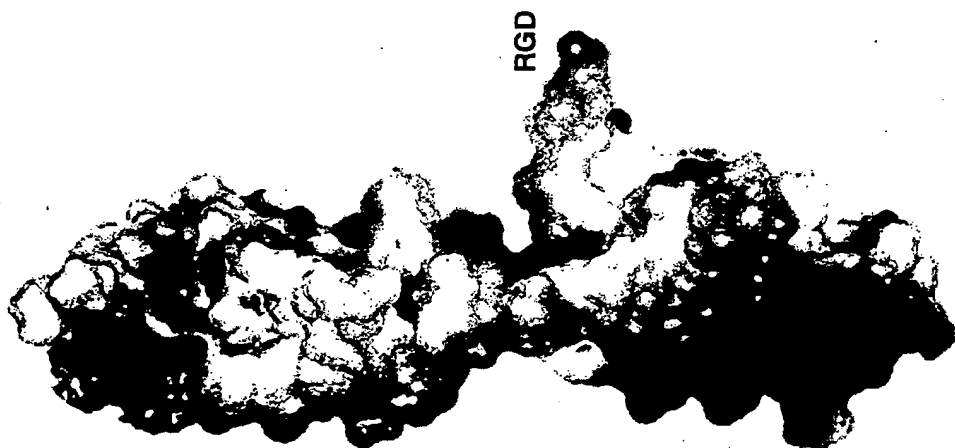


FIG. 5

6/25



95Fn3

10Fn3

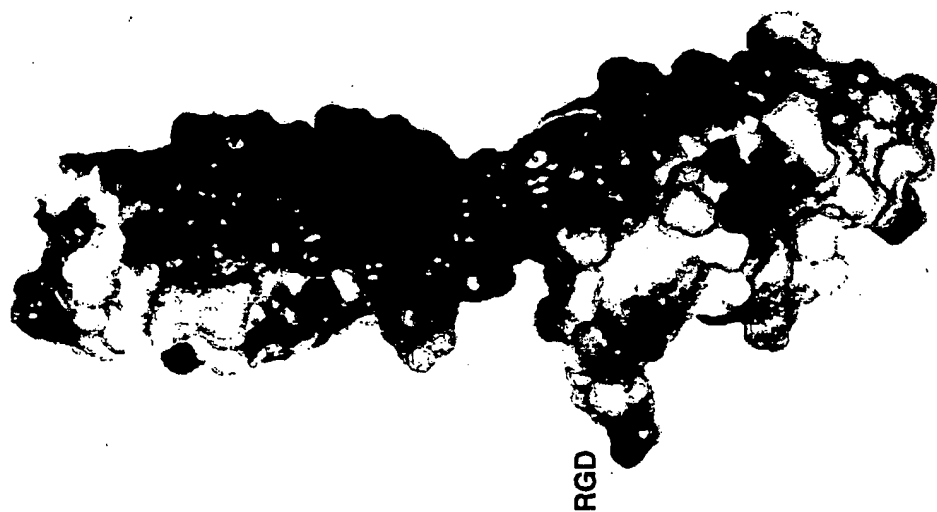


FIG. 6

7/25

$^7\text{Fn3}$

$^8\text{Fn3}$

$^9\text{Fn3}$

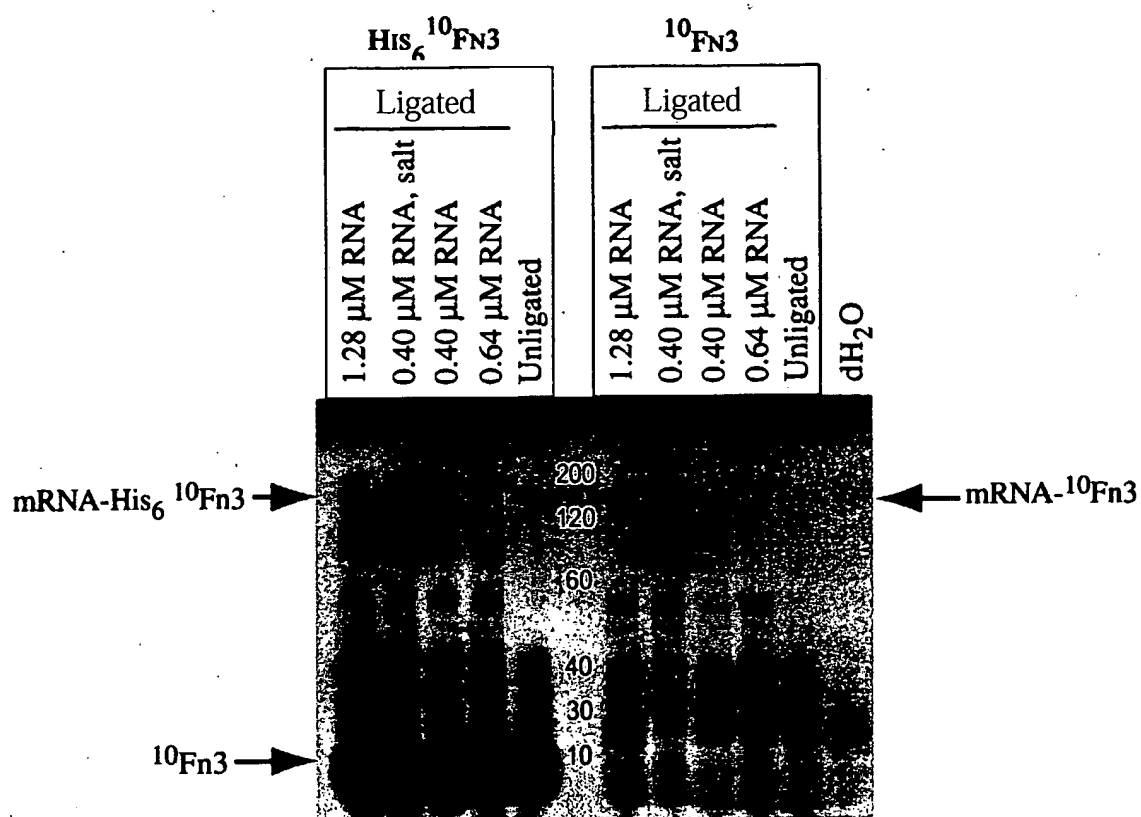
$^{10}\text{Fn3}$



FIG. 7

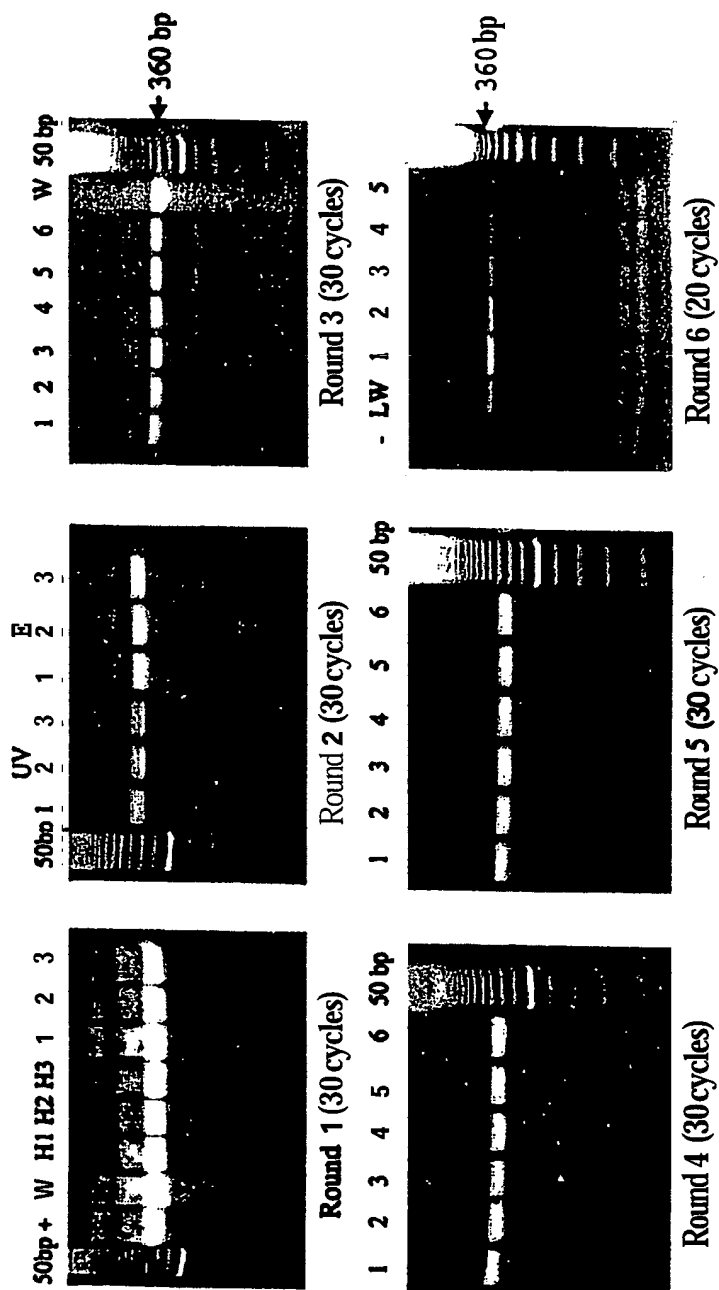
8/25

FIG. 8



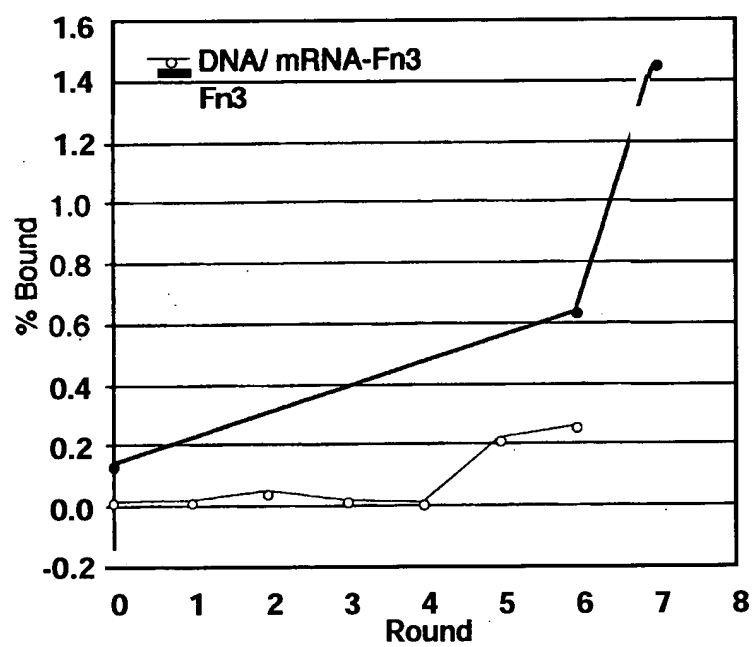
9/25

FIG. 9



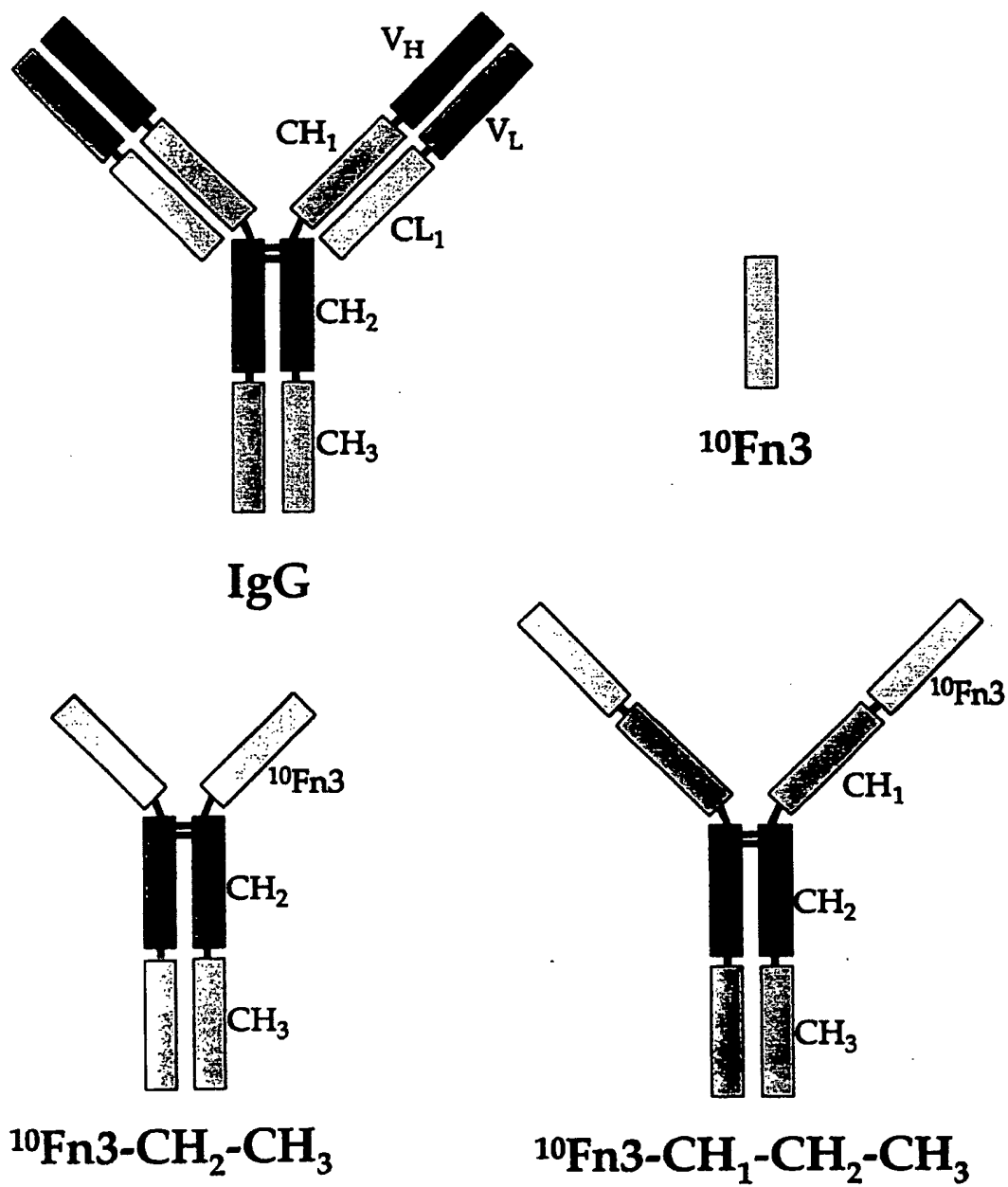
10/25

FIG. 10



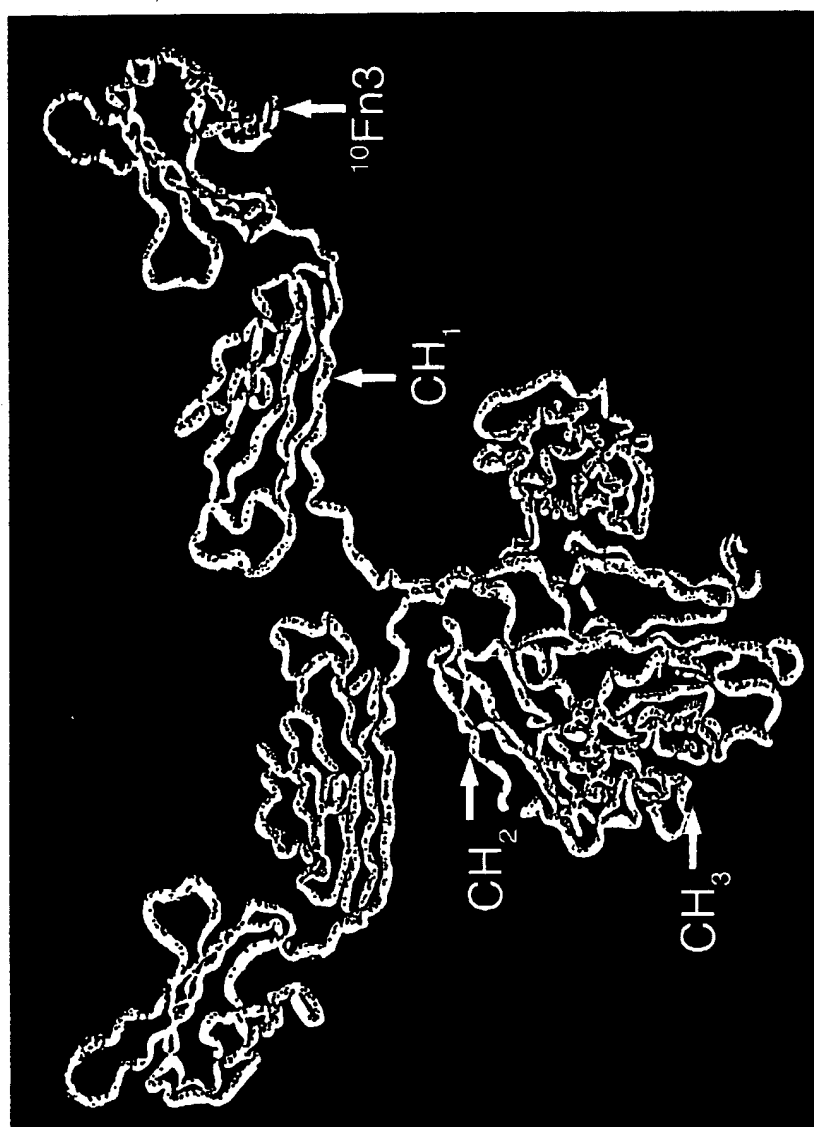
11/25

FIG. 11



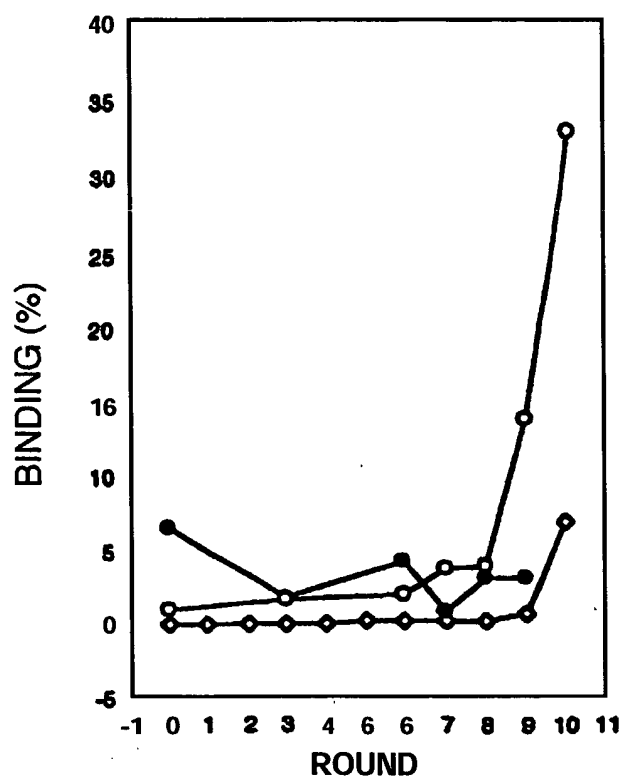
12/25

FIG.12



13/25

FIG. 13



14/25

FIG. 14

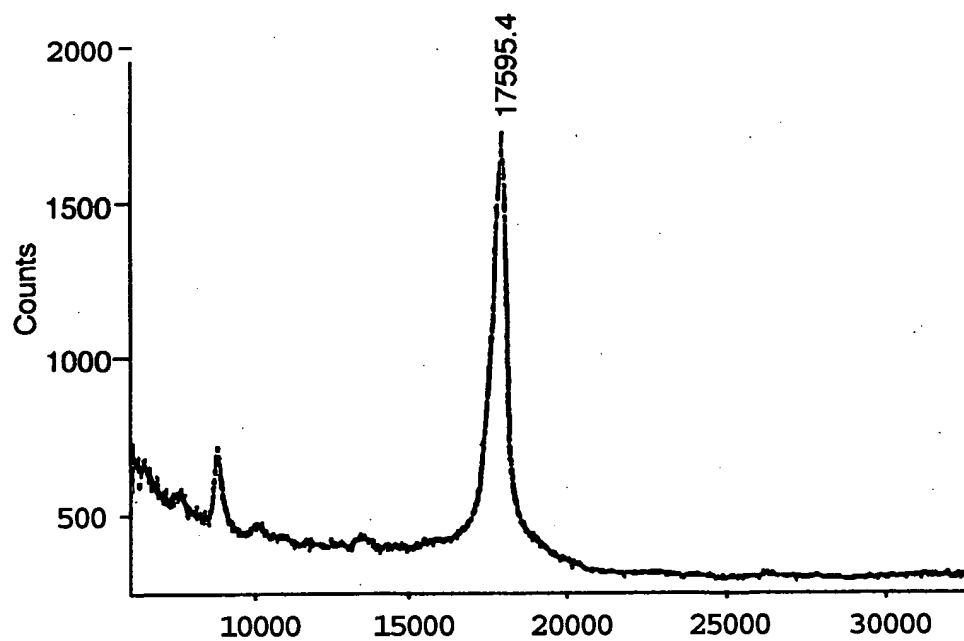
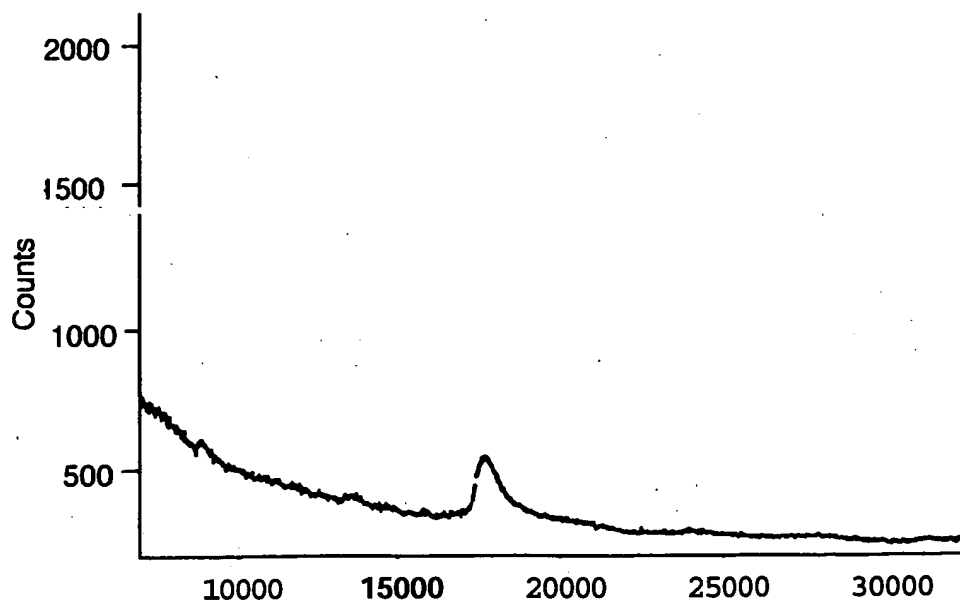


FIG. 15



15/25

FIG. 16

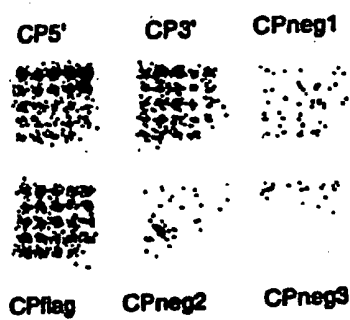


FIG. 17

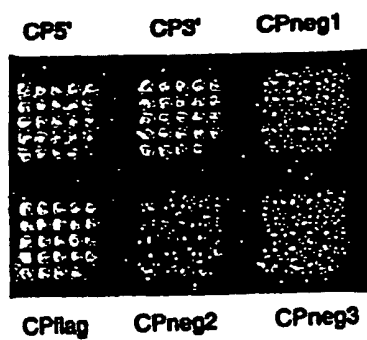


FIG. 18

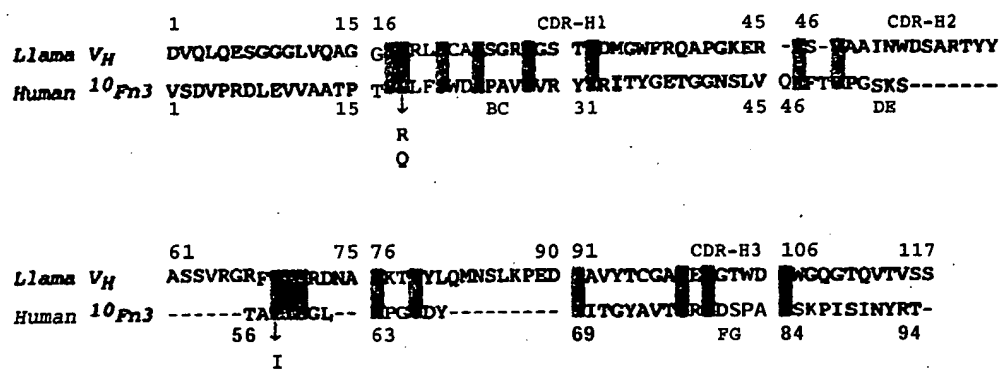
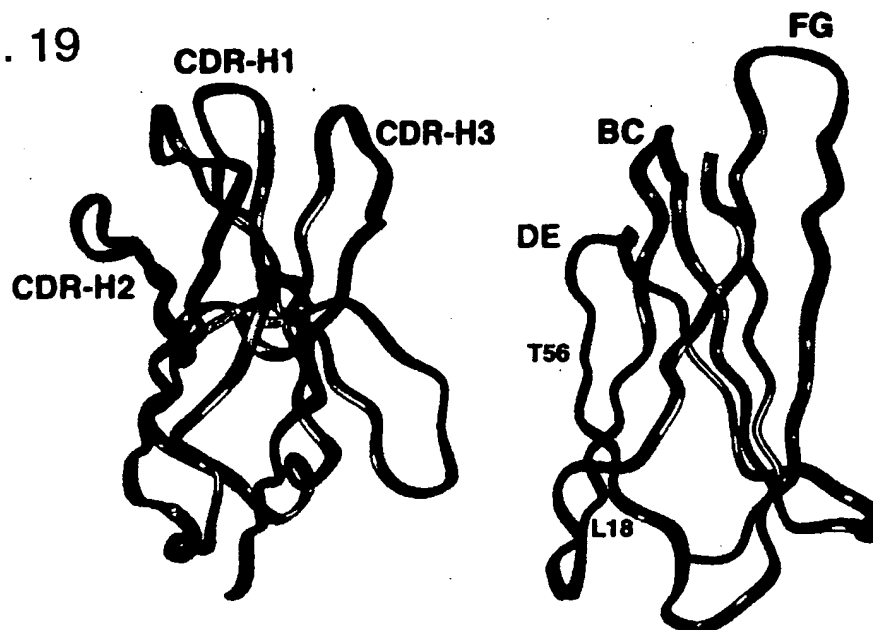
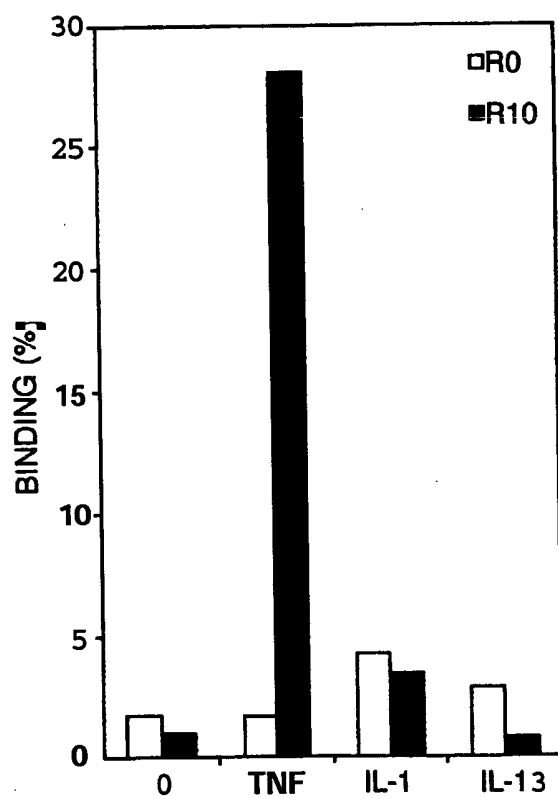


FIG. 19

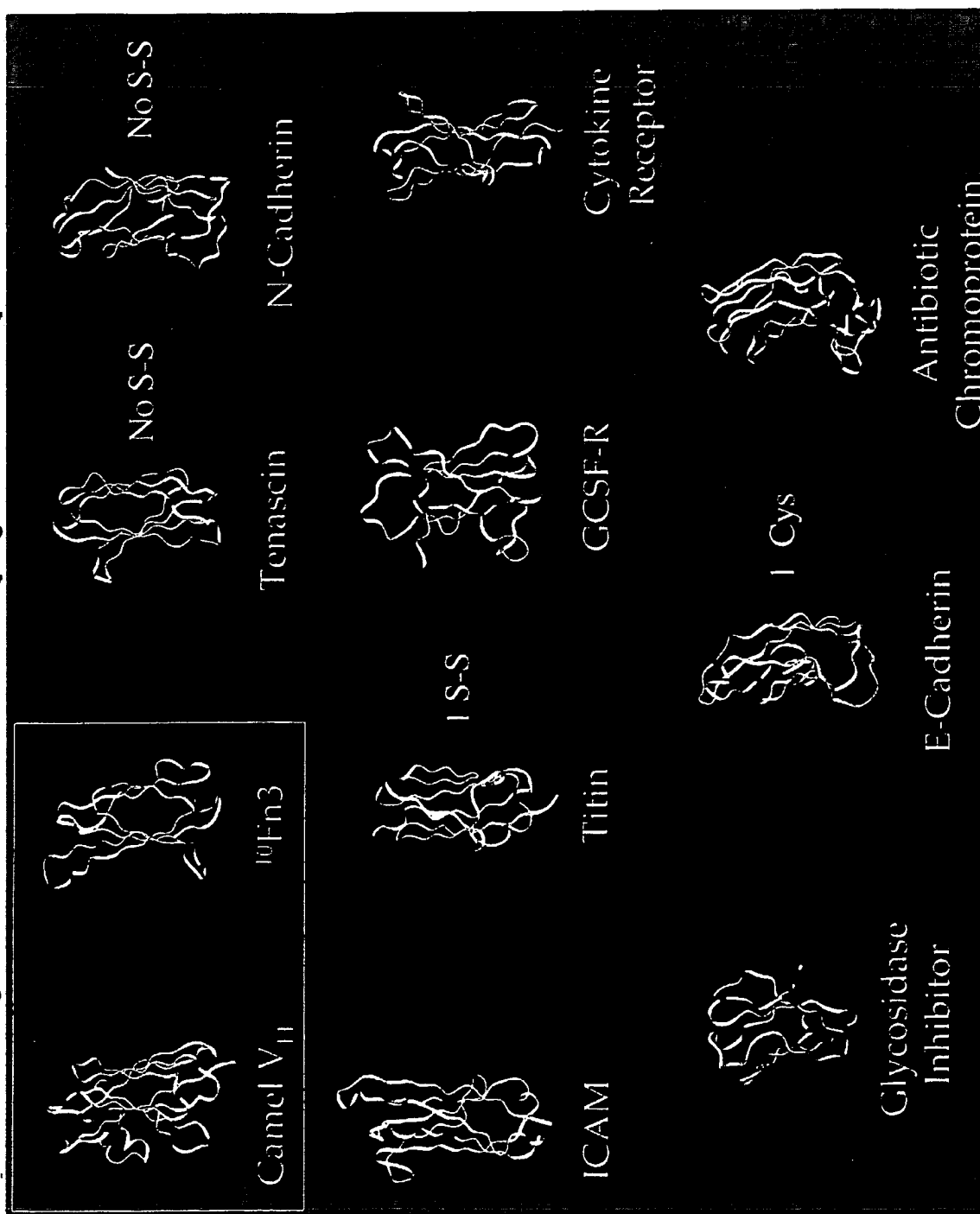


17/25

FIG. 20

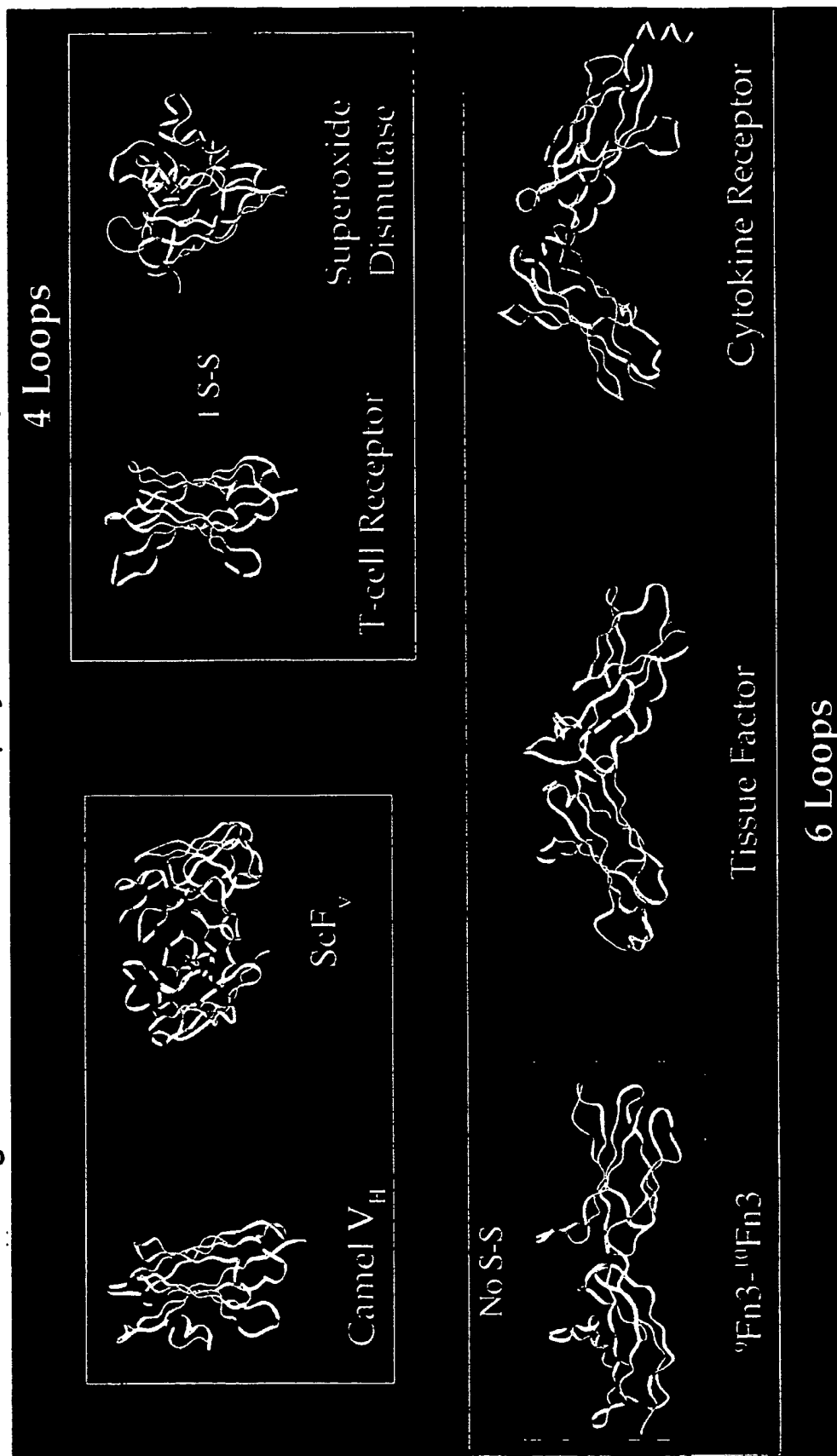
18/25

FIG. 21 IgG-like Scaffolds for the Display of Three Loops



19/25

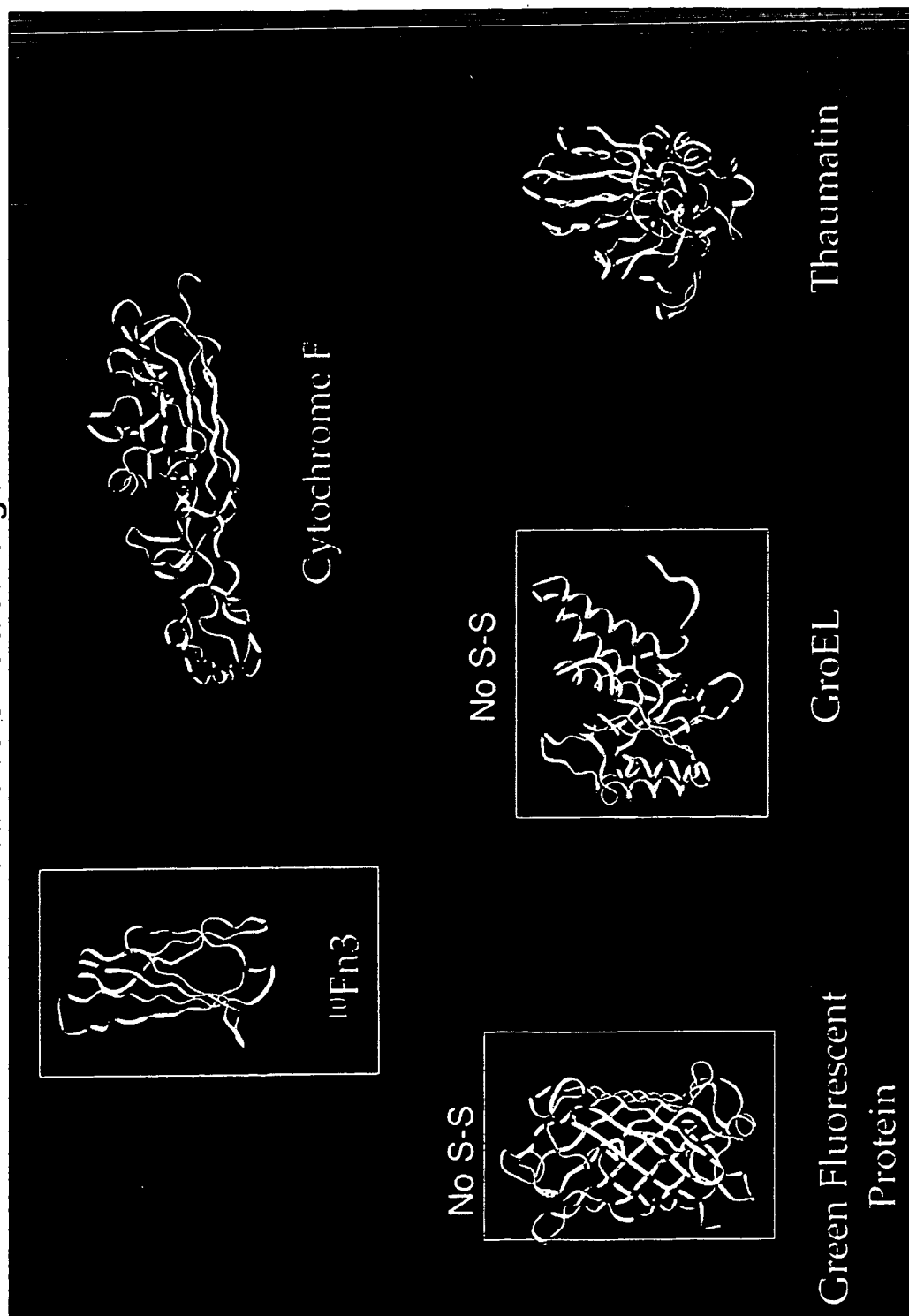
FIG. 22 IgG-like Scaffolds for the Display of Four or Six Loops



20/25

FIG. 23

Scaffolds Unrelated to IgG



21/25

FIG. 24A

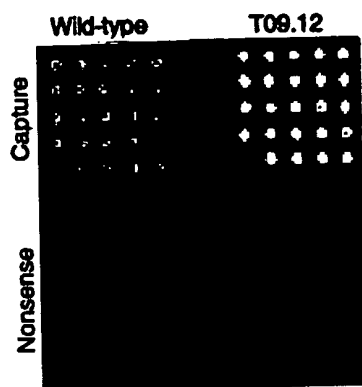


FIG. 24B

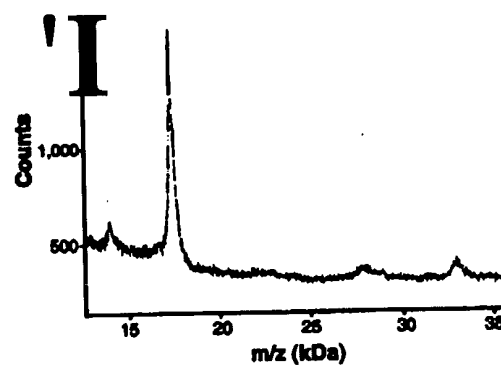


FIG. 24C

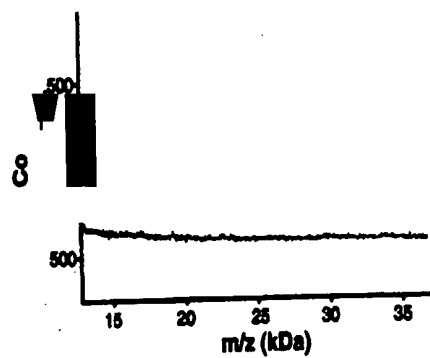


FIG. 24D

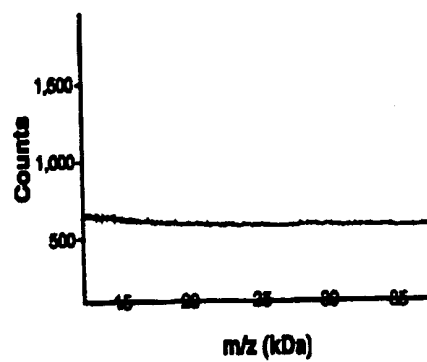


FIG. 25A

ROUND 9 (32)

ROUND 9 (32)				BC	DZ				FG				SEQ. ID. NO.
	1	15	16	DAPAVTV3031	45	46	GSKS	60	61	74	75	GRODSPASSK	90
WT	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	33				
T09.08	VSEIPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	34				
T09.26	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	35				
T09.20	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	36				
T09.25	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	37				
T09.30	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	38				
T09.02	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	39				
T09.34	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	40				
T09.03*	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	41				
T09.13	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	42				
M9.29	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	43				
T09.01	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	44				
T09.14	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	45				
T09.24	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	46				
T09.19	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	47				
T09.31	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	48				
T09.07	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	49				
T09.36	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	50				
T09.16	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	51				
T09.23	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	52				
T09.21	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	53				
T09.33*	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	54				
T09.10	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	55				
T09.28	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	56				
T09.15	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	57				
T09.09	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	58				
T09.35	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	59				
T09.06	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	60				
T09.11	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	61				
T09.12*	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	62				
T09.17	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	63				
T09.04	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	64				
T09.05	VSDVPRDLEVVAATP	TSLLISWDAPAVTVR	YRITYGETGNSPV	QFTVFGSKSTATIS	GLKPGVDY-TITVYA	VTGRDSPASSK	PIS	INVRT	65				

23/25

FIG. 25B

ROUND	10	(29)	BC	DE	FG	SEQ. ID. NO.
WT	1	15 16	DAPATV3031	45 46	GSKS	60 61
						74 75
						75GRGDSPASSK
						33
						INRT
						66
						INRT
						67
						INRT
						68
						INRT
						69
						INRT
						70
						INRT
						71
						INRT
						72
						INRT
						73
						INRT
						74
						INRT
						75
						INRT
						76
						INRT
						77
						INRT
						78
						INRT
						79
						INRT
						80
						INRT
						81
						INRT
						82
						INRT
						83
						INRT
						84
						INRT
						85
						INRT
						86
						INRT
						87
						INRT
						88
						INRT
						89
						INRT
						90
						INRT
						91
						INRT
						92
						INRT
						93
						INRT
						94
						INRT
						95
						INRT
						96
						INRT
						97
						INRT
						98
						INRT
						99
						INRT
						100
						INRT
						101
						INRT
						102
						INRT
						103
						INRT
						104
						INRT
						105
						INRT
						106
						INRT
						107
						INRT
						108
						INRT
						109
						INRT
						110
						INRT
						111
						INRT
						112
						INRT
						113
						INRT
						114
						INRT
						115
						INRT
						116
						INRT
						117
						INRT
						118
						INRT
						119
						INRT
						120
						INRT
						121
						INRT
						122
						INRT
						123
						INRT
						124
						INRT
						125
						INRT
						126
						INRT
						127
						INRT
						128
						INRT
						129
						INRT
						130
						INRT
						131
						INRT
						132
						INRT
						133
						INRT
						134
						INRT
						135
						INRT
						136
						INRT
						137
						INRT
						138
						INRT
						139
						INRT
						140
						INRT
						141
						INRT
						142
						INRT
						143
						INRT
						144
						INRT
						145
						INRT
						146
						INRT
						147
						INRT
						148
						INRT
						149
						INRT
						150
						INRT
						151
						INRT
						152
						INRT
						153
						INRT
						154
						INRT
						155
						INRT
						156
						INRT
						157
						INRT
						158
						INRT
						159
						INRT
						160
						INRT
						161
						INRT
						162
						INRT
						163
						INRT
						164
						INRT
						165
						INRT
						166
						INRT
						167
						INRT
						168
						INRT
						169
						INRT
						170
						INRT
						171
						INRT
						172
						INRT
						173
						INRT
						174
						INRT
						175
						INRT
						176
						INRT
						177
						INRT
						178
						INRT
						179
						INRT
						180
						INRT
						181
						INRT
						182
						INRT
						183
						INRT
						184
						INRT
						185
						INRT
						186
						INRT
						187
						INRT
						188
						INRT
						189
						INRT
						190
						INRT
						191
						INRT
						192
						INRT
						193
						INRT
						194
						INRT
						195
						INRT
						196
						INRT
						197
						INRT
						198
						INRT
						199
						INRT
						200
						INRT
						201
						INRT
						202
						INRT
						203
						INRT
						204
						INRT
						205
						INRT
						206
						INRT
						207
						INRT
						208
						INRT
						209
						INRT
						210
						INRT
						211
						INRT
						212
						INRT
						213
						INRT
						214
						INRT
						215
						INRT
						216
						INRT
						217
						INRT
						218
						INRT
						219
						INRT
						220
						INRT
						221
						INRT
						222
						INRT
						223
						INRT
						224
						INRT
						225

FIG. 25C

ROUND 14 (22)

	BC	DE	FG	SEQ. ID. NO.
WT	1 15 16 DAPAVTV3031	45 46 GSKS 60 61	75 76GRGDSPASSK 90	
T14.25	VSDVPRDLEVVAAATP TELLISMDAPAVTVR YRITYGETGNSPV QESTVPGSKSTATIS GLKPGVDY-TITVYA VTGRGDSPASSKPIIS			INVRT 33
T14.03	VSDVPRDLEVVAAATP TELLISMDTHNAYNG YRITYGETGNSPV REFTVPHPEVTATIS GLKPGVDD-TITVYA VTNHMPRLRIPGPIIS			INVRT 96
T10.06	VSDVPRDLEVVAAATP TELLISMDTRTNANTR YRITYGETGNSPV QESTAPNPPPTATIG GLKPGVDY-TITVYA VTFDGSRRHMLTKPIIS			INVRT 97
T14.12	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 81
T14.13	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 98
T14.17	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 99
T14.05	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 100
T14.14	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 101
T14.23	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 102
T14.26	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VADKSDTYKYDDPIIS			INVRT 103
T14.20	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 104
T14.19	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 105
T14.11	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 106
T14.10	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 107
T14.22	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 108
T14.21	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 109
T14.01	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 110
T14.06	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 111
T14.08	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 112
T14.02	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 113
T14.15	VSDVPRDLEVVAAATP TSRLISMRSGLOSR YRITYGETGNSPV QESTVPPWASLAIIS GLKPGVDY-TITVYA VTDKSDTYKYDDPIIS			INVRT 114
				INVRT 115
				INVRT 116

24/25

FIG. 25D

ROUND M12 (24)

SEQ. ID. NO.

FG

DE

BC

1	15 16	DAPAVTV3031	45 46	61	75	76GRGDSPASSK 90	33
WT	VSDVPRDLEVVAAATP	TSLLSWMDAPAVTVR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTGRGDSPASSKPI8	INVRT	117
M12.05	VSDVPRDLEVVAAATP	TSLLSWSPENDAHK	YRITYGEGNSPV	GLKPGVDY-TSVVYA	VTDQSYTYTYPIS	INVRT	118
M12.17	VSDVPRDLEVVAAATP	TSLLSWESQPTXGR	YRITYGEGNSPV	GRKPGADY-TITVYA	VTIKORIFLFGPI8	ISYRT	119
M12.12	VSDVPRDLEVVAAATP	TSLLSWESQPTXGR	YRITYGEGNSPV	GRKPGADY-TITVYA	VTIKORIFLFGPI8	ISYRT	120
M12.17	VSDVPRDLEVVAAATP	TSLLSWESQPTXGR	YRITYGEGNSPV	GRKPGADY-TITVYA	VTIKORIFLFGPI8	ISYRT	121
M12.06	VSDVPRDLEVVAAATP	TSLLSWDTSMAYNG	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTIHMPRLRIFGPI8	INVRT	122
M12.01	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDESIVKYDDPVS	INVRT	123
M12.01	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDESIVKYDDPVS	INVRT	124
M12.24	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSG	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTENVGRLDTRYPI8	IDCRT	125
M12.25	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTENVGRLDTRYPI8	INVRT	126
M12.23	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	XTSATPSRPNVHPIS	INVRT	127
M12.04	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTHERDT-----PI8	INVRT	128
M12.03	VSDVPRDLEVVAAATP	ISQLTSMQFPQNGSR	YRITYGEGNSPV	GLKPGVDY-AITVYA	ATHGKPPHHT---	INVRT	129
M12.07	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTENSTAOPEYPI8	FNRT	130
M12.13	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VAFPPGYFLTEMPIS	INVRT	131
M12.09	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	132
M12.16	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	133
M12.22	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	134
M12.26	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	135
M12.15	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	136
M12.18	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	137
M12.14	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	138
M12.21	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	139
M12.19	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	140
M12.08	VSDVPRDLEVVAAATP	TSLLSWNRSGI-QSR	YRITYGEGNSPV	GLKPGVDY-TITVYA	VTDSGTYRYDDPI8	INVRT	140

25/25